RUMLIG ØKONOMETRI
I TEORI OG PRAKSIS

9. OG 11. SEMESTERS PROJEKT

GRUPPE G3-113

INSTITUT FOR MATEMATISKE FAG
AALBORG UNIVERSITET

Abstract

This paper treats classical as well as Bayesian model theory and how to verify spatial econometric models. Furthermore we will assign great weight to the spatial autoregressive model SAR and go into particulars about its structure and parameters. Maximum likelihood estimation is explained and used for finding the parameters of the SAR model. Monte Carlo approximation is useful in the event of a huge quantity of data and in this particular case the method approximates parts of the SAR model. The estimated parameters of the SAR model are verified using statistical hypothesis tests, confidence intervals, the Wald test and residual plots. In contradistinction to the classical approach the Bayesian approach is explained and used for finding the parameters of the SAR model. The algorithm of Markov chain Monte Carlo is a part of the Bayesian approach and the algorithm includes Gibbs sampling and Metropolis Hastings sampling alike. Most of the theory is translated into practice using data from the real estate agent HOME.

Institut for Matematiske Fag
Aalborg Universitet
Fredrik Bajers Vej 7G
9220 Aalborg Ø
http://www.math.aau.dk

Titel: Rumlig økonometri
- i teori og praksis

Tema:

Rumlige, økonometriske modeller

Projektperiode:
Efterårssemestret 2012

Projektgruppe:
G3-113

Deltagere:

Helle Jakobsen

Lea Nørgreen Gustafsson

Vejleder:
Jakob Gulddahl Rasmussen

Oplagstal: 5

Sideantal: 124

Afsluttet den 21-12-2012

Synopsis:

Rapporten omhandler både klassisk og bayesiansk teori om modeller og modeltest inden for rumlige, økonometriske modeller med fokus på den rumligt afhængige SAR-model. Egenskaber ved SAR-modellen, dens opbygning og parametre bliver gennemgået i detaljer. Der er en beskrivelse af maximum likelihood-estimation generelt, og hvordan man ved hjælp af maximum likelihood-estimation kan finde udtryk for parametrene i SAR-modellen. Monte Carlo-approksimation behandles med henblik på tilfælde, hvor der haves store datasæt, således at approksimationer for led i SAR-modellen kræver særlig opmærksomhed. Der ses på modeltest med henblik på SAR-modellens parametre, herunder begreber som hypotesetest og konfidensintervaller, Waldtesten samt residualer. For at sammenligne den klassiske tilgang med den bayesianske tilgang til SAR-modellen introduceres bayesiansk teori og metode generelt, og der ses på den bayesianske fremgangsmåde og estimation af SAR-modellens parametre. Der er i den forbindelse fokus på Markov Chain Monte Carlo-estimation anvendt i form af Markov Chain Monte Carlo-algoritmen med Gibbs-sampling og Metropolis Hastings-sampling. Store dele af ovenstående teori bliver anvendt i praksis til analyse af et datasæt med boligpriser fra ejendomsmægleren HOME.

Forord

Denne projektrapport er udarbejdet i efteråret 2012 ved Institut for Matematiske Fag på Aalborg Universitet. Rapporten er udarbejdet i et samarbejde mellem kandidatstuderende på 9.-semester Helle Jakobsen og specialestuderende på 11.-semester Lea Nørgreen Gustafsson. I udarbejdelsen af rapporten er anvendt forskellige kilder, men den primære kilde er "Introduction to Spatial Econometrics" af James P. LeSage og Robert Kelley Pace, der sammen med de øvrige, anvendte kilder er i litteraturlisten.

Rapporten henvender sig til den interesserede læser, der forventes at have et vist kendskab til grundlæggende sandsynlighedsregning og have gennemført statistiske kurser svarende til universitetsniveau, så derfor vil basale, statistiske begreber ikke blive gennemgået i detaljer i rapporten.

Der rettes stor tak til vejleder Jakob Gulddahl Rasmussen for inspirerende vejledning og for konstruktiv kritik gennem hele projektforløbet.

Læsevejledning

Rapporten er delt op i 8 hoveddele: Indledning, SAR-modellen, Maximum likelihood-estimation for SAR-modellen, Monte Carlo-approksimation, Modeltest, Bayesianske, rumlige, økonometriske modeller og Databehandling.

I kapitel 1 indledes med en motivation for valg af emne, og der gives et kort historisk overblik over rumlige økonometriske modeller. Desuden vil der være et overblik over den anvendte metode og nogle overvejelser angående rumlighed.

Kapitel 2 introducerer SAR-modellen, dens opbygning, parametre og egenskaber, og gennemsnitlige påvirkninger for SAR-modellen vil også blive gennemgået. Der vil ligeledes blive gennemgået specificering af vægte til SAR-modellen.

I kapitel 3 beskrives maximum likelihood-estimation generelt, og hvordan man ved hjælp af maximum likelihood-estimation kan finde udtryk for parametrene i SAR-modellen.

I kapitel 4 behandles Monte Carlo-approksimation med henblik på tilfælde, hvor der haves store datasæt, således at approksimationer for led i SAR-modellen kræver særlig opmærksomhed. Desuden analyseres, hvilke konsekvenser en sådan Monte Carlo-approksimation har for en af de modelparametre, der har stor interesse i SAR-modellen.

Kapitel 5 introducerer modeltest med henblik på SAR-modellens parametre, herunder hypotesetest samt begreberne signifikansniveau, p-værdi og konfidensinterval. Herudover ses på blandt andet Waldtesten, for hvilken der gøres nogle praktiske overvejelser. Desuden indføres residualer, der ofte anvendes i forbindelse med modeltest.

Kapitel 6 omhandler en bayesiansk tilgangsvinkel til SAR-modellen. Den bayesianske metode vil blive gennemgået generelt, og der ses på den bayesianske fremgangsmåde til SAR-modellen og estimation af dens parametre. Der vil i den forbindelse være fokus på Markov Chain Monte Carlo-estimation anvendt i form af Markov Chain Monte Carlo-algoritmen med Gibbs-sampling og Metropolis-Hastings-sampling.

I Kapitel 7 vil store dele af ovenstående teori blive anvendt i praksis i programmet R til analyse af et datasæt indeholdende en række oplysninger om boligsalg fra ejendomsmægleren HOME.

Notation

Ved kildehenvisninger anvendes formen [forfatter(e), årstal (, placering i kilden)], eksempelvis [LeSage and Pace, 2009] eller [LeSage and Pace, 2009, kap. 2].

Definitioner, sætninger, lemmaer og eksempler er nummereret på formen X.Y, hvor X er det aktuelle kapitel, og Y er et fortløbende heltal startende ved 1 i hvert kapitel. Ligninger er nummereret på samme vis, men på formen (X.Y). Når der henvises til definitioner, sætninger, eksempler og kode fra R skrives for eksempel „sætning X.Y“, og når der henvises til ligninger skrives blot „(X.Y)“.

Om notationen bemærkes desuden følgende:

- Beviser afsluttes med ■, og eksempler afsluttes med □.
- Vektorer angives med fed skrift, eksempelvis $\mathbf{x}$, og matricer med stort bogstav som X.
- Modelparametre angives med græske bogstaver som α, β og ρ.

Indholdsfortegnelse

Indledning 1

1.1 Motivation

Netop fordi rumlig økonometri er en forholdsvis ny gren inden for økonometrien, og fordi det er et felt i stadig udvikling, synes vi det er spændende at se nærmere på teorien og metoden inden for dette felt og undersøge, hvad det kan bruges til i praksis. Det er spændende at se på en rumlig model, da den kan inddrage flere afhængige og uafhængige variable og muligvis derfor give mere præcise forecast end ikke-rumlige modeller.

1.2 Overblik

Dette afsnit er skrevet på baggrund af [Anselin, 2010].

Emnet rumlig økonometri er gået fra blot at være en marginal disciplin i 1960'erne og 1970'erne til at være en del af mainstream i dag. Rumlig økonometri er nu både et bredere kendt og mere velanset felt inden for anvendt økonometri, og teorien og metoden vinder større og større indpas. Det ses blandt andet gennem den stigende mængde af litteratur, der udgives om emnet, og ved at flere og flere eksperter inddrager rumlighed, når de arbejder med økonometri.

En af grundene til, at der er gået så lang tid, før rumlige modeller har vundet indpas i økonometriens verden, kan være, at der inden for feltet længe har været et fokus på geografisk modellering, som ikke umiddelbart har kunnet overføres direkte til teknikker til anvendt rumlig økonometri, og dermed har der været en form for begrænsning for anvendelse af teori og metode for rumlige modeller. Men nu er metoden og teorien blevet videreudviklet, således at flere og flere fagområder har muligheden for at anvende rumlighed inden for deres felt. Desuden kan der også ses en forbindelse mellem teknologiske forbedringer samt tilgængelighed af beregningsmetoder og den øgede anvendelse af rumlige modeller. Et skift i teoretiske perspektiver har også affødt interessen for at se nærmere på naboeffekter og rumlige spillovereffekter, hvilket vil sige interaktion mellem observationer.

Det felt, som rumlig økonometri beskæftiger sig med, defineres som „en delmængde af økonometriske metoder, der involverer rumlige aspekter, som er til stede i tværsnitsobservationer og observationer, der er defineret af både tid og rum. Variable relaterede til placering, distance og grupperinger behandles eksplicit i modelspecifikation, -estimation, -test og forecast" [Anselin, 2010, s. 5]. Tværsnitsobservationer opfattes som et endimensionalt datasæt, hvor der observeres mange objekter på ét specifikt tidspunkt.

Således er der fire, vigtige elementer, der definerer metodologien i den moderne, rumlige økonometri, nemlig modelspecifikation, modelestimation, modeltest og rumlig forecasting.

Modelspecifikation

Modelspecifikation drejer sig om det formelle, matematiske udtryk for rumlig afhængighed og rumlig ensartethed i regressionsmodeller til at beskrive data. Overordnet set kan rumlighed opdeles i to hovedformer for rumlige effekter, nemlig rumlig afhængighed og rumlig ensartethed.

Rumlig afhængighed kan ses som en form for afhængighed mellem tværsnitsobservationer i den forstand, at korrelationsstrukturen eller kovariansstrukturen mellem stokastiske variable ved forskellige placeringer er udledt af en specifik gruppering bestemt af distance eller rumlig gruppering af observationerne i det geografiske rum. Teknikkerne til at modellere rumlig afhængighed er ikke blot en udvidelse af teknikker fra autoregressive processer til to dimensioner, men er mere komplicerede end som så. Rumlig afhængighed i modeller ses typisk i form af inddragelsen af rumligt laggede variable, altså vægtede gennemsnit af værdier for naboliggende observationer for en givet observation. Naborelationer specificeres oftest ved en rumlig vægtmatrix. Rumligt laggede variable kan inkluderes for den afhængige variabel (rumlige lag-modeller), baggrundsvariable (rumlige tværsnitsmodeller), og fejlled (rumlig fejlmodel), og desuden kan der haves kombinationer af disse, hvilket resulterer i et rigt udvalg af detaljerede og fyldestgørende, rumlige modeller til beskrivelse af datasæt.

Rumlig ensartethed er et tilfælde af observeret eller ikke observeret ensartethed. Det antages ofte, at naboliggende observationer påvirker hinanden, således at disse tilnærmelsesvist ensrettes, samt at udefrakommende påvirkninger giver tilsvarende udsving i naboliggende observationer, hvilket kan udnyttes i modelspecifikationen. Denne ensartethed kan enten være kendt og dermed definerbar i modellen, eller det kan være en ukendt faktor, som er til stede i data uden at indgå direkte som en parameter i modellen. I modsætning til rumlig afhængighed kræves der ikke altid separate metoder til at behandle ensartethed. Det eneste rumlige aspekt ved ensartethed er den yderligere information, der måske stilles til rådighed ved brug af rumlige strukturer. For eksempel kan dette måske resultere i modeller med rumligt varierende koefficienter og rumlige, strukturelle ændringer. Rumlig ensartethed kommer matematisk til udtryk på flere måder. Ensartethed kan klassificeres som enten diskret ensartethed eller kontinuert ensartethed. Diskret ensartethed består af en forudbestemt mængde af rumligt adskilte enheder, inden for hvilke modelparametre må variere. Kontinuert ensartethed specificerer, hvordan regressionsparametrene varierer i rummet, enten ved at følge en forudbestemt form eller ved at være bestemt ud fra data gennem en estimationsproces. Inden for statistik specificeres den rumlige ensartethed som et specialtilfælde af variation i stokastiske koefficienter.

Modelestimering

Når de rumlige effekter er inddraget og specificeret i en regressionsmodel, anvendes estimeringsmetoder, som tager højde for den simultanitet, der følger af rumlig afhængighed. En vigtig metode til dette er maksimum likelihood-estimation, der bruges til at finde værdier for de regressionsparametre, som er af interesse i modellen.

Modeltest

Ved modeltest er der fokus på at finde afvigelser mellem data og standard, rumlige modelspecifikationer, der kan indikere et behov for rumlige alternativer. Sådanne test kan for eksempel være baseret på maksimum likelihood-estimation som f.eks. likelihood ratio-test eller Waldtesten.

Rumlig forecasting

Rumlig forecasting er ikke så anvendt en disciplin som de tre foregående inden for økonometri. Rumlig forecasting består af metoder til forudsigelse af ukendte, stokastiske størrelser i modeller, der inddrager rumlig afhængighed.

1.3 Rumlige data

Dette afsnit er skrevet på baggrund af [Wall, 2003].

Når der haves en mængde observationer, der spreder sig ud over et geografisk område, som for eksempel huspriser i forskellige kommuneer, er det ikke altid muligt at inddrage alle de variable, der kan have påvirkning på observationerne. Det antages dog som regel, at de udeladte variable i de forskellige regioner ligner hinanden, da for eksempel huspriser i et område ofte er påvirket af huspriser i de nærliggende områder, og dermed bliver fejlleddene rumligt autokorrelerede. Observationer kan måles i gitre, det vil sige forbundne, diskrete punkter i rummet, da for eksempel huspriser knytter sig til bestemte, geografiske punkter, hvor husene ligger, og der findes to måder at modellere sådanne gitterdata på. Begge metoder er specialtilfælde af den generelle, rumlige proces $\{Z(s)|s \in D\}$, hvor forskellen ses i antagelserne om mængden D.

Den første metode antager, at D er en kontinuert mængde, hvor observationerne stammer fra. Ved brug af et Voronoi-diagram kan der summeres over alle observationerne, som knyttes til et geografisk punkt midt i hvert område, for eksempel midt i en kommune, og afstanden mellem de forskellige centrer bruges til at definere om to områder er naboer. Samme afstande benyttes også til at danne en rumlig kovariansstruktur via en funktion, der beskriver graden af rumlig afhængighed mellem stokastiske processer. Denne funktion kan for eksempel være en funktion af par af positioner, så kovariansen afhænger af afstanden mellem observationerne. Det antages ofte, at kovariansen aftager des længere observationerne er fra hinanden. Et problem ved metoden er for det første, at det er forsimplet at placere alle observationerne i centrum af områderne, og for det andet at det ikke er praktisk muligt at måle data kontinuert, hvilket gør, at en kontinuert model kun vil være en tilnærmelse til de diskrete data. Dog kan man betragte diskrete data som kontinuerte, hvis de er observeret tæt på hinanden.

Den anden metode vil blive anvendt i kapitel 2 og tager udgangspunkt i, at D antages at være en diskret mængde. Her defineres naboer til at være de områder, hvis grænser rører hinanden, og således kan der bestemmes en autoregressiv model. To af de modeller, der tager højde for det diskrete mønster, er SAR-modellen (Spatial Autoregressiv Model) og CAR-modellen (Conditionally Autoregressive Model). Oprindeligt blev modellerne

udviklet som modeller for endelige, regulære gitre, og dermed er de i familie med den velkendte, stationære, autoregressive proces. Effekten fra naboliggende strukturer og rumlige korrelationsparametre på kovariansstrukturen er dog ikke forstået i detaljer, når SAR- og CAR-modellerne anvendes på irregulære gitre, hvilket vil sige, hvor observationerne ikke har samme antal naboer, og observationerne ikke nødvendigvis har samme afstand til hinanden. Hvis modellerne derimod anvendes på et irregulært, endeligt gitter, kan modellerne antyde kovariansens struktur, som angiver, at der ikke er konstant varians for regioner, der er samme antal naboer fra hinanden, og at der heller ikke er lige kovarians mellem samme. I rapporten har vi valgt kun at se på SAR-modellen.

En anden metode at beskrive rumlige data på er gennem bayesianske modeller, der tager højde for forhåndsviden om modellens parametre.

SAR-modellen 2

Dette kapitel er skrevet på baggrund af [LeSage and Pace, 2009, kapitel 2], [Wall, 2003] og [Horvath and Johnston].

Når der arbejdes med data som for eksempel priser på huse set i forhold til beliggenhed, kan hver observation fra datasættet opfattes som et diskret punkt i en mængde D.

Definition 2.1 (Gitter):
Datamængden $D = \{A_1, \ldots, A_n\}$ kaldes et gitter, hvis elementerne i $\{Z(A_i)|A_i \in (A_1, \ldots, A_n)\}$ er normalfordelte, og $\{A_1, \ldots, A_n\}$ opfylder, at

$$A_1 \cup A_2 \cup \ldots \cup A_n = D$$
$$A_i \cap A_j = 0 \quad \forall \quad j \neq i.$$

Et gitter er således en mængde observationer, der kan opdeles i disjunkte delmængder.

2.1 SAR-modellen

SAR-modellen er en rumlig, økonometrisk model, som beskriver data, der følger en rumlig, autoregressiv proces. Den kan også skrives på form som en datagenererende proces, hvorfra der kan simuleres nye data, som kan sammenlignes med de oprindelige data for at se, hvor god modellen er.

Definition 2.2 (DGP - Data Generating Process):
Den datagenererende proces er en model, der er speciferet således, at nye data kan simuleres derfra, idet den afhængige variabel er isoleret.

Der findes mange forskellige datagenererende processer, og det er relevant at finde den datagenererende proces, der bedst beskriver de data, der arbejdes med. En sådan proces kan både inkludere baggrundsvariable og rumlig afhængighed i forklarende såvel som afhængige variable. I praksis anvendes den fundne model til forecasting. Desuden kan man ved at se på SAR-modellens datagenererende proces lettere aflæse dataenes fordeling, hvorimod man ved modellens grundform nemmere kan aflæse og analysere hver parameters betydning i modellen.

Den datagenererende proces for SAR-modellen, som vi vender tilbage til (se sætning 2.7), er en omskrivning af selve modellen $Z(A_i)$, der kan skrives som

$$\begin{aligned} Z(A_i) &= \mu_i + \sum_{j=1}^{n} \rho w_{ij}(Z(A_j) - \mu_j) + \varepsilon_i \\ \varepsilon_i &\sim N(0,\sigma^2) \\ E[Z(A_i)] &= \mu_i \quad \text{for} \quad i = 1,\ldots,n. \end{aligned} \tag{2.1}$$

Her er ρw_{ij} kendte eller ukendte konstanter, og $w_{ii} = 0$ for $i = 1,\ldots,n$. Modellen kaldes simultan, idet fejlleddet ε_i generelt er korreleret med $\{Z(A_i)|j \neq i\}$. I det tilfælde at n er endelig, kan der defineres en matrix W for alle elementerne w_{ij}, så $W = (w_{ij})$. Mængden $\mathbf{Z} = \begin{bmatrix} Z(A_1) & Z(A_2) & \ldots & Z(A_n) \end{bmatrix}^T$ kan også skrives som vektoren $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^T$, der er observeret data, og de er multivariat normalfordelte (se appendiks B.1) som

$$\begin{aligned} \mathbf{Y} &\sim N_n\left(\boldsymbol{\mu},\Sigma\right) \\ \boldsymbol{\mu} &= \begin{bmatrix} \mu_1 & \mu_2 & \ldots & \mu_n \end{bmatrix}^T. \end{aligned}$$

Formen af det anvendte gitter D angiver strukturen for matricen W.

Definition 2.3 (Nabomatrix W):
Matricen W består af elementerne (w_{ij}), der opfylder

$$\begin{aligned} w_{ij} &= 1 \quad \text{hvis } A_i \text{ har en fælles grænse med } A_j \\ w_{ij} &= 0 \quad \text{hvis } i = j \\ w_{ij} &= 0 \quad \text{ellers,} \end{aligned}$$

hvor A_i og A_j er delmængder af datamængden $D = \{A_1,\ldots,A_n\}$.

Vægtmatricen W kan altid konstrueres således, at den er rækkestokastisk, hvilket vil sige, at rækkerne summerer til en. Det kan gøres ved hjælp af en invertibel matrix som for eksempel en permutationsmatrix. I de fleste tilfælde er det nemmest, hvis matricen W er rækkestokastisk, hvilket kan gøres ved at sætte $W = (w^*_{ij})$, hvor

$$\begin{aligned} w^*_{ij} &= \frac{w_{ij}}{w_{i+}} \\ w_{i+} &= \sum_{k=1}^{n} w_{ik}, \end{aligned}$$

hvor w_{i+} angiver, at vi summerer over alle de ettaller, der er i j'te række i W. Det gør, at W bliver rækkestokastisk.

Der ses nu på et eksempel på konstruktion af W.

Eksempel 2.4:
Eksemplet er skrevet på baggrund af [LeSage and Pace, 2009, s. 8-10].

Der tages udgangspunkt i de fem regioner i Danmark, hvilket vil sige Region Nordjylland, Region Midtjylland, Region Syddanmark, Region Sjælland og Region Hovedstaden, der ses på figur 2.1.

Figur 2.1: Regioner i Danmark fra [A/S].

Det er tydeligt, at Region Nordjylland er nabo til Region Midtjylland og omvendt, og at Region Midtjylland er nabo til Region Syddanmark og omvendt osv. Disse kaldes førsteordensnaboer. Desuden er det klart, at Region Nordjylland kun har én nabo givet ved Region Midtjylland, og at Region Midtjylland har to naboer givet ved Region Nordjylland og Region Syddanmark. Vi kan lave en (5×5)-vægtmatrix W, der beskriver naborelationer mellem de fem regioner, givet ved

$$W = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & 0 & 1 & 0 & 0 & 0 \\ Mi & 1 & 0 & 1 & 0 & 0 \\ Sy & 0 & 1 & 0 & 1 & 0 \\ Sj & 0 & 0 & 1 & 0 & 1 \\ Ho & 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Der står et 1-tal på den (i,j)'te plads, hvis regionerne repræsenteret ved disse pladser er naboer, og hvis ikke noteres et 0. For eksempel er Region Sjælland den eneste førsteordennabo til Region Hovedstaden, hvilket både kan ses ved at kigge på rækken eller søjlen ud for Region Hovedstaden. Desuden bemærkes det, at alle diagonalindgange er 0, hvilket udtrykker, at ingen regioner anses for at være deres egen nabo. Idet vi ønsker at se på rumlig afhængighed og linearkombinationer af værdier for naboobservationer, kan vi gøre W rækkestokatisk. Dette gøres ved at dividere alle indgange med antallet af

indgange, hvori der står et 1-tal, således at W er givet ved vægtmatricen

$$W = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & 0 & 1 & 0 & 0 & 0 \\ Mi & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ Sy & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ Sj & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ Ho & 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{2.2}$$

□

Vi vender nu tilbage til udtrykket i (2.1) og omskrive det ved først at sætte $y_i = Z(A_i)$, hvilket giver

$$y_i = \mu_i + \sum_{j=1}^{n} \rho w_{ij}(y_j - \mu_j) + \varepsilon_i.$$

Leddet $\sum_{j=1}^{n} w_{ij}(y_j - \mu_j)$ kan skrives som en vektor for alle værdier $i = 1,\ldots,n$, så

$$\begin{bmatrix} \sum_{j=1}^{n} w_{1j}(y_j - \mu_j) \\ \vdots \\ \sum_{j=1}^{n} w_{nj}(y_j - \mu_j) \end{bmatrix} = \begin{bmatrix} w_{11} & \ldots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \ldots & w_{nn} \end{bmatrix} \begin{bmatrix} y_1 - \mu_1 \\ \vdots \\ y_n - \mu_n \end{bmatrix} = W(\mathbf{y} - \boldsymbol{\mu}).$$

Det gør, at (2.1) kan skrives på matrixform som

$$\begin{aligned} \mathbf{y} &= \boldsymbol{\mu} + \rho W(\mathbf{y} - \boldsymbol{\mu}) + \boldsymbol{\varepsilon} \\ &= \boldsymbol{\mu} + \rho W\mathbf{y} - \rho W\boldsymbol{\mu} + \boldsymbol{\varepsilon} \\ &= \rho W\mathbf{y} + (I_n - \rho W)\boldsymbol{\mu} + \boldsymbol{\varepsilon}. \end{aligned}$$

Sættes $\boldsymbol{\mu} = (I_n - \rho W)^{-1} X\boldsymbol{\beta}$, hvor X er en matrix, og $\boldsymbol{\beta}$ er en vektor, fås

$$\begin{aligned} \mathbf{y} &= \rho W\mathbf{y} + (I_n - \rho W)(I_n - \rho W)^{-1} X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \end{aligned} \tag{2.3}$$

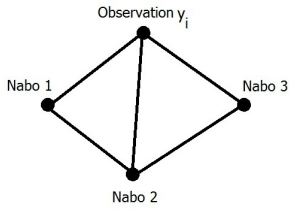
Skrivemåden (2.3) er anvendt i definition 2.5 og vil blive anvendt i resten af rapporten.

Definition 2.5 (SAR-modellen - Spatial Autoregressiv Model):
Den rumlige, autoregressive model, SAR-modellen, er givet ved

$$\begin{aligned} \mathbf{y} &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N_n(\mathbf{0}, \sigma^2 I_n), \end{aligned}$$

hvor X er en designmatrix med baggrundsvariable, og $\boldsymbol{\beta}$ er en parametervektor. Fejlvektoren $\boldsymbol{\varepsilon}$ er fordelt efter den multivariate normalfordeling med n dimensioner med en $(n \times 1)$-nulvektor som middelværdi og $\sigma^2 I_n$ som variansmatrix. Matricen W er en rumlig $(n \times n)$-vægtmatrix, der beskriver rumlig afhængighed mellem naboobservationer, og ρ er en parameter, der beskriver styrken af rumlig afhængighed mellem observationerne.

Matricen W har nuller på diagonalen og er rækkestokastisk, hvilket vil sige, at rækkerne summerer til 1. Faktoren $W\mathbf{y}$ i definition 2.5 kaldes den rumlige lagvektor, og hvert element i vektoren $W\mathbf{y}$ repræsenterer linearkombinationer af naboliggende observationer. For at lave en model, der passer til de pågældende data, estimeres regressionsparametrene $\boldsymbol{\beta}$ og σ samt parameteren ρ. Det ses, at ρ i definition 2.5 sammen med W beskriver den reelle afhængighed mellem alle observationer. Hvis $\mathbf{y}$ er normalfordelt, giver ρ information om naborelationer, hvor $\rho > 0$ angiver, at der er positiv korrelation mellem naboliggende observationer, hvilket vil sige, at naboliggende observationer påvirker og ligner hinanden. Tilfældet $\rho < 0$ angiver, at der typisk er negativ korrelation mellem naboobservationer, men i tilfælde hvor naborelationerne er komplicerede, kan det være sværere at fortolke betydningen af $\rho < 0$. For eksempel kan en observation y_i have tre naboer, hvor to af dem begge er nabo til den tredje, som vist på figur 2.1.



Figur 2.2: En observation y_i med naboer, hvor fortolkning af ρ ikke er entydig.

Hvis observation y_i har negativ korrelation med nabo 1, kan y_i ikke også have negativ korrelation med nabo 2, hvis naboerne indbyrdes også er negativt korrelerede. I sådanne tilfælde kræves der en dybere analyse af ρ's betydning for naborelationerne. Hvis $\rho = 0$ er naboobservationer ukorrelerede, og idet y_i er normalfordelt, haves det, at naboobservationer er uafhængige af hinanden.

Eksempel 2.6:

Eksemplet er skrevet på baggrund af [LeSage and Pace, 2009, s. 8-10 og s. 14-15].

Vi ser nu på en forsimplet udgave af SAR-modellen givet ved

$$\mathbf{y} = \rho W\mathbf{y} + \boldsymbol{\varepsilon}$$
$$\boldsymbol{\varepsilon} \sim N_n\left(\mathbf{0}, \sigma^2 I_n\right).$$

Udtrykket beskriver en sammenhæng mellem vektoren $\mathbf{y}$ og vektoren $W\mathbf{y}$, der repræsenterer en linearkombination af naboværdier til hver observation. Vi multiplicerer hermed (5×5)-vægtmatricen W i (2.2) fra eksempel 2.4 med en (5×1)-vektor $\mathbf{y}$, der repræsenterer værdier for hver region, hvormed der fås en rumlig lagvektor $W\mathbf{y}$ af den afhængige, variable vektor $\mathbf{y}$. Vektoren $W\mathbf{y}$ er en (5×1)-vektor, der repræsenterer værdierne af den rumlige lagvektor for hver observation i for $i = 1, \ldots, 5$. Der fås en rumlig lagvektor bestående af simple gennemsnit af værdier fra naboobservationer til hver region. Lagvektoren

$W\mathbf{y}$ er givet ved

$$W\mathbf{y} = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & 0 & 1 & 0 & 0 & 0 \\ Mi & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ Sy & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ Sj & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ Ho & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} y_2 \\ \frac{y_1+y_3}{2} \\ \frac{y_2+y_4}{2} \\ \frac{y_3+y_5}{2} \\ y_4 \end{bmatrix}.$$

Der ses nu på potenser af vægtmatricen W, og der tages stadig udgangspunkt i de fem regioner i Danmark. Hvor vætmatricen i første potens, altså blot W, repræsenterer førsteordensnaboer, så vil W i anden potens, W^2, repræsentere andenordensnaboer, mens W^3 er tredjeordenanboer osv. En andenordensnabo er defineret som en førsteordensnabos nabo. Således er Region Nordjyllands eneste andenordensnabo givet ved RegionSyddanmark, mens Region Syddanmark har to andenordensnaboer givet ved Region Nordjylland og Region Hovedstaden. Desuden er for eksempel Region Nordjylland andenordensnabo til sig selv, ligesom alle de andre regioner også er det, idet alle regioner er nabo til deres egen nabo, hvilket netop er definitionen på en andenordensnabo. Idet alle observationer er andenordensnabo til sig selv, vil der på diagonalen i W^2 være positive elementer, hvis en observation har mindst én nabo og ikke blot nuller som i W, der repræsenterer førsteordensnaboer. Vægtmatricen W^2 for de fem regioner er givet ved

$$W^2 = \begin{bmatrix} & No & Mi & Sy & Sj & Ho \\ No & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ Mi & 0 & \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ Sy & \frac{1}{4} & 0 & \frac{1}{2} & 0 & \frac{1}{4} \\ Sj & 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ Ho & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix},$$

hvor det ses, at andenordensnaboerne for Region Nordjylland er Region Nordjylland selv og Region Syddanmark, ligesom Region Syddanmarks andenordensnaboer er Region Nordjylland, Region Syddanmark og Region Hovedstaden. Andenordensnaboerne tilskrives mindre betydning end førsteordensnaboerne, hvilket er en generel tendens, så jo højere orden naboer har, jo mindre indflydelse har naborelationerne på en pågældende observation. Det afhænger selvfølgelig også af værdien for ρ.

Senere skal vi se på udtryk, hvor det udnyttes, at

$$\left(I_n - \rho W\right)^{-1} = \rho^0 W^0 + \rho^1 W^1 + \rho^2 W^2 + \rho^3 W^3 + \ldots = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 + \ldots.$$

Der kræves her udregning af matricen for andenordensnaboer og de højereordens naboer. □

Vi kan også skrive SAR-modellen op for en enkelt observation, så

$$y_i = \sum_{j=1}^{n} (X\boldsymbol{\beta})_{ij} + \rho \sum_{j=1}^{n} W_{ij} y_j + \varepsilon_i$$
$$\varepsilon_i \sim N(0,\sigma^2) \quad \text{for} \quad i = 1,\ldots,n.$$

I praksis kan man tilføje et led, $\alpha\iota_n$, som er en konstant søjlevektor af n ettaller multipliceret med parameteren α, der sammen har til formål at tage højde for den situation, hvor $\mathbf{y}$ ikke har middelværdien nul. Leddet $\alpha\iota_n$ er i definition 2.5 indeholdt i vektoren $X\boldsymbol{\beta}$, men hvis modellen for eksempel omskrives til at indeholde to $\boldsymbol{\beta}$-vektorer, $\boldsymbol{\beta}_1$ og $\boldsymbol{\beta}_2$, og to designmatricer, X_1 og X_2, i stedet for én, ville $\alpha\iota_n$ indgå to gange, så derfor isoleres leddet i visse tilfælde fra starten. SAR-modellen har da formen

$$\begin{aligned}\mathbf{y} &= \rho W\mathbf{y} + \alpha\iota_n + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim N_n(\mathbf{0},\sigma^2 I_n).\end{aligned}$$

Hvis vi omskriver udtrykket fra definition 2.5, fås den rumlige, autoregressive, datagenererende proces for SAR-modellen, der er vist i sætning 2.7.

Sætning 2.7 (SAR-modellen - den datagenererende proces):
Den rumlige, autoregressive, datagenererende proces for SAR-modellen er givet ved

$$\begin{aligned}\mathbf{y} &= (I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})\\ \boldsymbol{\varepsilon} &\sim N_n(\mathbf{0},\sigma^2 I_n),\end{aligned}$$

hvor $\mathbf{y}$ er en vektor af n observationer, W er en rækkestokastisk vægtmatrix, X er en designmatrix over baggrundsvariable, og ρ og $\boldsymbol{\beta}$ er parametre. Vektoren $\boldsymbol{\varepsilon}$ er en multivariat normalfordelt fejlvektor.

Bevis:
Der foretages omskrivningen

$$\begin{aligned}\mathbf{y} &= \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} && \Rightarrow\\ (I_n - \rho W)\mathbf{y} &= X\boldsymbol{\beta} + \boldsymbol{\varepsilon} && \Rightarrow\\ \mathbf{y} &= (I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})\\ \boldsymbol{\varepsilon} &\sim N_n(\mathbf{0},\sigma^2 I_n).\end{aligned}$$

Her er det antaget, at matricen $I_n - \rho W$ er invertibel, hvilket bevises senere i rapporten ved lemma 2.12. ■

Sætning 2.8:
Værdierne for datavektoren $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}^T$ er i SAR-modellen fordelt som

$$\mathbf{Y} \sim N_n\left(\left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}, (I_n - \rho W)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T\right).$$

Bevis:
Middelværdien for $\mathbf{y}$ i den datagenererende proces for SAR-modellen er

$$\begin{aligned}\mathrm{E}\left[\mathbf{y}\right] &= \mathrm{E}\left[(I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})\right]\\ &= \mathrm{E}\left[(I_n - \rho W)^{-1}X\boldsymbol{\beta} + (I_n - \rho W)^{-1}\boldsymbol{\varepsilon}\right]\\ &= \mathrm{E}\left[(I_n - \rho W)^{-1}X\boldsymbol{\beta}\right] + \mathrm{E}\left[(I_n - \rho W)^{-1}\boldsymbol{\varepsilon}\right]\\ &= \mathrm{E}\left[(I_n - \rho W)^{-1}X\boldsymbol{\beta}\right]\\ &= (I_n - \rho W)^{-1}X\boldsymbol{\beta},\end{aligned}$$

da $\boldsymbol{\varepsilon}$ er multivariat normalfordelt med middelværdi nul, og $(I_n - \rho W)^{-1} X\boldsymbol{\beta}$ er et tal.

Variansen for $\mathbf{y}$ i den datagenererende proces for SAR-modellen er

$$\begin{aligned}\text{Var}\left[\mathbf{y}\right] &= \text{Var}\left[(I_n - \rho W)^{-1} X\boldsymbol{\beta} + (I_n - \rho W)^{-1}\boldsymbol{\varepsilon}\right] \\ &= \text{Var}\left[(I_n - \rho W)^{-1}\boldsymbol{\varepsilon}\right] \\ &= (I_n - \rho W)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T,\end{aligned}$$

da $(I_n - \rho W)^{-1} X\boldsymbol{\beta}$ er et tal og variansen for $\boldsymbol{\varepsilon}$ er $\sigma^2 I_n$. ■

Det, der er kendetegnet ved SAR-modellen, er, at den medtager afhængighed mellem observationerne fra naboliggende punkter udtrykt ved vektoren $\rho W\mathbf{y}$, det vil sige rumlig afhængighed i den afhængige variabel $\mathbf{y}$. SAR-modellen anvendes derfor ved antagelse om rumlig ensartethed i data.

Eksempel 2.9:

Eksemplet er skrevet på baggrund af [Statistik, Opslag: Folketal], [Krak] og [LeSage and Pace, 2009, s. 2-3 og 17-20].

For at rejse fra Aalborg til København med tog eller bil passerer man gennem alle fem regioner i Danmark. Regionerne kan opfattes som liggende på en linje, da man fra Region Nordjylland kommer til Region Midtjylland og derefter videre til Region Syddanmark og så videre, indtil man ender i Region Hovedstaden. Rejsetiderne gennem hver region er cirka

Rejsetider i minutter	**Område**
36	Region Nordjylland (Aalborg - Hobro)
80	Region Midtjylland (Hobro-Vejle)
84	Region Syddanmark (Vejle-Korsør)
60	Region Sjælland (Korsør- Roskilde)
34	Region Hovedstaden (Roskilde-København K)

Rejsetiden fra hver af regionerne til Region Hovedstaden kan opskrives som en vektor $\mathbf{y}$ givet ved

$$\mathbf{y} = \begin{bmatrix} 294 \\ 258 \\ 178 \\ 94 \\ 34 \end{bmatrix},$$

hvor tiden er målt i minutter, og den første værdi tilhører Region Nordjylland, mens den sidste tilhører Region Hovedstaden. Da Region Hovedstaden er et større område, tager det også tid at rejse fra et sted til et andet deri, så denne værdi er ikke nul. Designmatricen X består af baggrundsvariable, som rejsetiden $\mathbf{y}$ afhænger af. Det kan for eksempel være befolkningsantallet i hvert område, da det har indflydelse på trafikken og dermed rejsetiden, ligesom den afstand, der tilbagelægges i hver region, også har indflydelse på

rejsetiden. Designmatricen X kan derfor have formen

$$X = \begin{matrix} \text{Afstand} & \text{Befolkningsantal} \\ \end{matrix} \begin{bmatrix} 425{,}2 & 543.116 \\ 371{,}5 & 1.159.189 \\ 242{,}1 & 1.093.729 \\ 111{,}9 & 757.257 \\ 37{,}2 & 1.442.536 \end{bmatrix},$$

hvor afstandene er målt i kilometer. Rejsetiden fra Region Nordjylland til Region Hovedstaden, y_1, afhænger af rejsetiden fra Region Midtjylland til Hovedstaden, y_2, ligesom den afhænger af resten af rejsetiderne, så vi kan anvende SAR-modellen fra definition 2.5, hvor W er givet i (2.2). Parametrerne $\hat{\rho}$ og $\hat{\boldsymbol{\beta}}$ kan estimeres ved hjælp af maximum likelihood, der gennemgås i kapitel 3, ved brug af R, og estimaterne er

$$\hat{\rho} = -0{,}55$$
$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 0{,}91 \\ 0{,}0000024 \end{bmatrix}.$$

Fra definition 2.5 haves det, at en forudsigelse fra SAR-modellen med de estimerede parametre har formen

$$\hat{\mathbf{y}}^{(1)} = (I_n - \hat{\rho} W)^{-1} X \hat{\boldsymbol{\beta}}.$$

SAR-modellen modellerer spillover-effekter, det vil sige, hvis der sker en ændring i i'te række i matricen X, vil det påvirke alle indgange i $\mathbf{y}$ og ikke kun y_i. Spill-overeffekten kan for eksempel ses ved at ændre i befolkningsantallet for en af regionerne og derefter se, hvilke ændringer der sker i rejsetiderne $\mathbf{y}$. Det kan eksempelvis antages, at vandstanden stiger på grund af global opvarmning, og 500.000 hollændere immigrerer til Region Syddanmark på grund af oversvømmelser, så den nye matrix X^* bliver

$$X^* = \begin{bmatrix} 425{,}2 & 543.116 \\ 371{,}5 & 1.159.189 \\ 242{,}1 & \mathbf{1.593.729} \\ 111{,}9 & 757.257 \\ 37{,}2 & 1.442.536 \end{bmatrix}.$$

Forudsigelsen

$$\hat{\mathbf{y}}^{(2)} = (I_n - \hat{\rho} W)^{-1} X^* \hat{\boldsymbol{\beta}}$$

anvender således matricen X^* med en ændring i befolkningsantallet i Region Syddanmark. I tabel 2.1 ses estimaterne af rejsetiden for begge tilfælde og forskellen mellem dem. Det er tydeligt, at en ændring i en enkelt værdi i matricen X påvirker rejsetiderne for alle regionerne.

Region	$\hat{\mathbf{y}}^{(1)}$	$\hat{\mathbf{y}}^{(2)}$	$\hat{\mathbf{y}}^{(2)} - \hat{\mathbf{y}}^{(1)}$
Nordjylland	261,9	262,1	0,26
Midtjylland	229,7	229,3	-0,47
Syddanmark	142,2	143,6	1,46
Sjælland	64,0	63,5	-0,47
Hovedstaden	2,1	2,4	0,26

Tabel 2.1: Estimater for rejsetider ved brug af SAR-modellen.

De totale spillover-effekter fås ved at summere over fjerde kolonne i tabel 2.1. Den direkte effekt aflæses ud for den observation, der er blevet ændret, mens de indirekte fås ved at trække den direkte effekt fra den totale spillover-effekt.

At to af regionerne får kortere rejsetid ved en ændring i en anden region er ikke helt, hvad man vil forvente, og det skyldes, at vi her ser på en meget forsimplet udgave af virkeligheden, hvor der kan være betydningsfulde baggrundsvariable, der er ikke er medregnet i modellen. Der kan rettes op på det ved at tilføje andre værdier i X, men da det ikke er formålet med dette eksempel, vil det ikke blive gjort her. □

2.2 Tidsafhængig motivation for SAR

Der findes en dynamisk motivation for SAR-modellen, idet den kan opstå som et resultat af tidsafhængighed for eksempel af beslutninger, der tages ved objekter i rummet, når disse beslutninger afhænger af tidligere naboliggende beslutninger. Den fundamentale forskel på almindelige, auteregressive processer og rumlige, autoregressive processer ses ved rumlig interaktion. Rumlig interaktion ses gennem naborelationer defineret af matricen ρW, gennem hvilken der kan forekomme feedback i form af spillover-effekt, det vil sige observation y_i kan påvirke observation y_j, men y_j kan også påvirke y_i. Der kan således ikke være tale om en direkte sammenligning med den tidsafhængige, autoregressive model, da SAR-modellen generelt ikke er en tidsafhængig model, men det kan i visse tilfælde være en fordel at opfatte den som en sådan.

Sætning 2.10:
SAR-modellen kan udtrykkes som en langsigtet ligevægtstilstand givet ved

$$\lim_{t\to\infty} \mathrm{E}[\mathbf{y}_t] = \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}.$$

hvor $\mathbf{y}_t$ er en vektor af observationer til tiden t, matricen ρW angiver naborelationer, designmatricen X indeholder baggrundsvariable, og $\boldsymbol{\beta}$ er en parametervektor.

Bevis:
Vi ser på den afhængige, variable vektor til tiden t, $\mathbf{y}_t$. Vektoren bestemmes ved brug af et rumligt, autoregressivt system, som afhænger af den afhængige variabel selv, hvilket sker gennem naboobservationer, hvor værdierne i den afhængige variabel er laggede i både tid og rum. Således fås et tids-lag for et vægtet gennemsnit af naboværdier for

den afhængige variabel observeret i den tidligere periode, $W\mathbf{y}_{t-1}$. I modellen inkluderes også nutidige karakteristika for den pågældende observation X_t, hvor $X_t = X$, hvis regionale karakteristika kan antages at være forholdsvist fastholdte over tid. Vi får hermed modellen

$$\mathbf{y}_t = \rho W\mathbf{y}_{t-1} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t.$$

Her bestemmes $\mathbf{y}_t$ ved kendskab til $\mathbf{y}_{t-1}$, og $\mathbf{y}_{t-1}$ bestemmes ved kendskab til $\mathbf{y}_{t-2}$ osv. Det vil sige, der haves en markovkæde. Vi kan dermed erstatte $\mathbf{y}_{t-1}$ med $\rho W\mathbf{y}_{t-2} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t-1}$, så

$$\begin{aligned}\mathbf{y}_t &= \rho W\left(\rho W\mathbf{y}_{t-2} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{t-1}\right) + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \\ &= X\boldsymbol{\beta} + \rho WX\boldsymbol{\beta} + \rho^2 W^2\mathbf{y}_{t-2} + \boldsymbol{\varepsilon}_t + \rho W\boldsymbol{\varepsilon}_{t-1}.\end{aligned}$$

Derfor kan $\mathbf{y}_t$ udtrykkes som en funktion af $\mathbf{y}_{t-s}$, hvor s er det antal tidsperioder, der springes tilbage i tiden, så det for $s < t$ fås, at

$$\begin{aligned}\mathbf{y}_t &= \left(I_n + \rho W + \rho^2 W^2 + \ldots + \rho^{s-1} W^{s-1}\right) X\boldsymbol{\beta} + \rho^s W^s\mathbf{y}_{t-s} + u \\ u &= \boldsymbol{\varepsilon}_t + \rho W\boldsymbol{\varepsilon}_{t-1} + \rho^2 W^2\boldsymbol{\varepsilon}_{t-2} + \ldots + \rho^{s-1} W^{s-1}\boldsymbol{\varepsilon}_{t-(s-1)} \\ &\Downarrow \\ \mathbf{y}_t &= \left(\sum_{j=0}^{s-1} \rho^j W^j\right) X\boldsymbol{\beta} + \rho^s W^s\mathbf{y}_{t-s} + \sum_{j=0}^{s-1} \rho^j W^j\boldsymbol{\varepsilon}_{t-j}.\end{aligned} \tag{2.4}$$

Det ses, at $\mathrm{E}[\boldsymbol{\varepsilon}_{t-r}] = 0 \Rightarrow \mathrm{E}[u] = 0$, for $r = 0, \ldots, s-1$. Desuden bliver betydningen af $\rho^s W^s\mathbf{y}_{t-s}$ lille for stort s under den antagelse, at $|\rho| < 1$, da W er rækkestokatisk, hvilket er ensbetydende med, at den har en numerisk største egenværdi på 1. Således kan vi betragte tværsnitsobservationerne som udfaldet eller den forventede værdi af en langsigtet ligevægtstilstand, hvilket vil sige

$$\lim_{t\to\infty} \mathrm{E}[\mathbf{y}_t] = \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta}.$$

Summen i (2.4) er her omskrevet til $\left(I_n - \rho W\right)^{-1}$ ved hjælp af den geometriske sumformel, idet der er konvergens, da W er rækkestokastisk, og det antages, at $|\rho| < 1$. ■

Nu vil vi se på, hvad der motiverer antagelsen $|\rho| < 1$, og hvorfor den er nødvendig.

Sætning 2.11:
Parameteren ρ skal opfylde $|\rho| < 1$, for at

$$\mathbf{y}_t = \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1} \rho^j W^j\right) X\boldsymbol{\beta} + \rho^s W^s\mathbf{y}_{t-s} + \sum_{j=0}^{s-1} \rho^j W^j\boldsymbol{\varepsilon}_{t-j}\right)$$

konvergerer.

Bevis:
Vi tager udgangspunkt i den autoregressive proces med rumlige lags fra (2.4). Grænseværdien for $\mathbf{y}_t$ kan findes ved at lade $s \to \infty$, hvilket er udtrykt ved

$$\mathbf{y}_t = \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1} \rho^j W^j\right) X\boldsymbol{\beta} + \rho^s W^s\mathbf{y}_{t-s} + \sum_{j=0}^{s-1} \rho^j W^j\boldsymbol{\varepsilon}_{t-j}\right). \tag{2.5}$$

Der er tre tilfælde, nemlig $|\rho| > 1$, $|\rho| = 1$ og $|\rho| < 1$. I alle tre tilfælde ser vi på udtrykket

$$\rho W = \rho V \Lambda_W V^{-1} \Rightarrow \rho^s W^s = \underbrace{\left(\rho V \Lambda_W V^{-1}\right)\left(\rho V \Lambda_W V^{-1}\right)\cdots\left(\rho V \Lambda_W V^{-1}\right)}_{s} = \rho^s V \Lambda_W^s V^{-1}, \tag{2.6}$$

jævnfør spektral dekomposition, og da W antages at være diagonaliserbar, og $VV^{-1} = I_n$. Da Λ_W er en diagonalmatrix haves, at

$$(\rho\Lambda_W)^s = \begin{bmatrix} (\rho\lambda_1)^s & 0 & \cdots & 0 \\ 0 & (\rho\lambda_2)^s & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & (\rho\lambda_n)^s \end{bmatrix}. \tag{2.7}$$

Hvis der eksisterer et i, så $|\rho\lambda_i| > 1$, svarer det til $|\rho| > 1$, da $\lambda_{\max} = 1$. Det ses dermed, at hvis $|\rho| > 1$, så vil ρ^s og Λ_W^s begge vokse, når s bliver større, og dermed vil $\rho^s W^s$ blive uendelig stor for stigende s. Matricen (2.7) konvergerer derfor ikke i dette tilfælde. Det ses hermed, at hvis $|\rho| > 1$ i (2.5), så har udtrykket ikke nogen grænseværdi, og systemet vil så at sige eksplodere eller i hvert fald aldrig stabilisere sig. Desuden bemærkes det, at observationer længere og længere væk (W^s) og påvirkninger ($\mathbf{y}_{t-s}$) samt fejl ($\boldsymbol{\varepsilon}_{t-(s-1)}$) langt tilbage i tiden ($t-s$) får større og større indflydelse, hvilket rent intuitivt ikke giver mening.

Hvis der for et i gælder, at $|\rho\lambda_i| = 1$, så er $|\rho| = 1$, da $\lambda_{\max} = 1$. Det resulterer i, at (2.7) konvergerer mod en matrix, der ikke er nul-matricen. Desuden ses det, at for $|\rho| = 1$, så reduceres (2.5) til

$$\mathbf{y}_t = \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1} W^j\right) X\boldsymbol{\beta} + W^s\mathbf{y}_{t-s} + \sum_{j=0}^{s-1} W^j\boldsymbol{\varepsilon}_{t-j}\right),$$

og dermed ses det efter samme princip som i (2.6), at processen vil have uendelig hukommelse, idet påvirkninger fra observationer langt tilbage i tiden ikke formindskes, men bliver ved med at have ligeså stor betydning for $\mathbf{y}_t$, som påvirkninger tættere på har.

Det ses, at hvis alle $\rho\lambda_i \in (-1,1)$, så konvergerer (2.7) mod nul-matricen. Da $\lambda_{\max} = 1$ vil det betyde, at $\rho \in (-1,1)$. Hermed haves det, at hvis (2.5) skal stabilisere sig, må der nødvendigvis gælde, at $|\rho| < 1$, hvilket kaldes den stationære antagelse, og systemet er dermed stationært. Vi betragter nu (2.5) og ser, hvad der sker med hvert enkelt led, når det antages, at $|\rho| < 1$.

Det første led kan skrives om ved hjælp af den geometriske sumformel, idet W er rækkestokastisk og det antages, at $|\rho| < 1$, så

$$\lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1}\rho^j W^j\right) X\boldsymbol{\beta}\right) = \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta},$$

det vil sige et konstant led. Da $|\rho| < 1$, så haves, at

$$\rho^s W^s \mathbf{y}_{t-s} \to 0 \quad \text{når} \quad s \to \infty.$$

Vi kan ikke sige med sikkerhed, hvordan sidste led i (2.5) opfører sig, når $|\rho| < 1$, men hvis leddet antages at konvergere, kan vi nu skrive

$$\begin{aligned}\mathbf{y}_t &= \lim_{s\to\infty}\left(\left(\sum_{j=0}^{s-1}\rho^j W^j\right)X\boldsymbol{\beta} + \rho^s W^s \mathbf{y}_{t-s} + \sum_{j=0}^{s-1}\rho^j W^j \boldsymbol{\varepsilon}_{t-j}\right)\\ &= \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + 0 + \sum_{j=0}^{\infty}\rho^j W^j \boldsymbol{\varepsilon}_{t-j}\\ &= \left(I_n - \rho W\right)^{-1} X\boldsymbol{\beta} + \sum_{j=0}^{\infty}\rho^j W^j \boldsymbol{\varepsilon}_{t-j}.\end{aligned}$$

Vi har således, at $\mathbf{y}_t$ kan skrives som en konstant adderet med en uendelig sum af påvirkninger fra fortiden, hvor fejlene har mindre og mindre betydning for $\mathbf{y}_t$, jo længere tilbage i tiden de er. Det samme gælder for naborelationerne jo længere væk de er. Værdien af ρ bestemmer således systemets hukommelse. ■

SAR-modellen kan ses som værende en form for grænsemodel for den almindelige, autoregressive proces. Det vil sige, at hvis vi betragter den autoregressive proces i det uendelige, haves udtrykket for SAR-modellen. Derfor gælder overvejelserne om ρ i den autoregressive proces fra dette afsnit også for ρ i SAR-modellen.

2.3 Kovariansmatricen for SAR-modellen

Det kan vises, at der er en tæt sammenhæng mellem den rækkestokastiske vægtmatrix W og dens egenværdier, et passende interval for parameteren ρ og invertibiliteten af matricen $I_n - \rho W$. Sammenhængen kommer til udtryk ved at undersøge udtrykket for kovariansmatricen i SAR-modellen nærmere. Kovariansmatricen er jævnfør sætning 2.8 givet ved

$$(I_n - \rho W)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T. \tag{2.8}$$

For en kovariansmatrix skal der gælde, at den er positivt definit og symmetrisk. Vi kan egentlig nøjes med at have en semidefinit kovariansmatrix, idet definitionen af normalfordelingen også er gældende for en sådan, men da vi er interesserede i en veldefineret tæthed for modellen, og vi ønsker at kovariansmatricen er invertibel, skærper vi kravene til kovariansmatricen til, at den skal være positivt definit. Vi vil bevise, at disse egenskaber under visse omstændigheder gælder for (2.8) ved først at bevise to lemmaer.

Lemma 2.12:
Matricen $I_n - \rho W$ er invertibel, hvis $\rho \in (\lambda_{\min}^{-1}, 1)$, hvor ρ er en parameter, W er en rækkestokastisk vægtmatrix med reelle egenværdier, og $\lambda_{\min}$ er den mindste egenværdi for W.

Bevis:
Matricen $I_n - \rho W$ er invertibel, hvis $|I_n - \rho W| \neq 0$. Der ses nu på matricen W. Enhver $(n \times n)$-matrix W kan opskrives på jordan kanonisk form (se [Lawrance, 2001, s. 39-40]),

så

$$J = V^{-1}WV \Rightarrow W = VJV^{-1}.$$

Hvis W er diagonaliserbar, er J blot en diagonalmatrix, og hvis ikke haves en øvre diagonal matrix, så

$$J = \begin{bmatrix} \lambda_1 & j_{1,2} & 0 & \dots & 0 \\ 0 & \lambda_2 & j_{2,3} & \ddots & \vdots \\ 0 & 0 & \lambda_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & j_{n-1,n} \\ 0 & \dots & 0 & 0 & \lambda_n \end{bmatrix},$$

hvor λ_i er W's egenværdier, og hvor $(J)_{ij}$ er 0 eller 1 for $i = 1,\dots,n-1$ og $j = i+1$. Det vil sige, at der haves en superdiagonal. Determinanten kan omskrives ved hjælp af ovenstående, så

$$|I_n - \rho W| = |I_n - \rho VJV^{-1}| = |V||I_n - \rho J||V^{-1}| = |I_n - \rho J|,$$

da $|V||V^{-1}| = |I_n| = 1$. Idet J har W's egenværdier λ_i for $i = 1,\dots,n$ på diagonalen, haves det, at

$$|I_n - \rho J| = \begin{vmatrix} 1-\rho\lambda_1 & \rho j_{1,2} & 0 & \dots & 0 \\ 0 & 1-\rho\lambda_2 & \rho j_{2,3} & \ddots & \vdots \\ 0 & 0 & 1-\rho\lambda_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \rho j_{n-1,n} \\ 0 & \dots & 0 & 0 & 1-\rho\lambda_n \end{vmatrix} = \prod_{i=1}^{n}(1-\rho\lambda_i),$$

da determinanten for en øvre, triangulær matrix er givet ved produktet af diagonalelementerne. Det ses, at

$$|I_n - \rho J| = \prod_{i=1}^{n}(1-\rho\lambda_i) \neq 0 \quad \text{hvis} \quad \rho \neq \frac{1}{\lambda_i}. \tag{2.9}$$

Vi ser nu nærmere på egenværdierne for W, der er antaget at være reelle. Da W er rækkestokastisk, haves det, at

$$W \cdot \iota_n = 1 \cdot \iota_n, \tag{2.10}$$

hvilket betyder, at $\lambda_k = 1$ er en egenværdi for W.

Idet $\text{tr}(W) = \sum_{i=1}^{n}\lambda_i$, og W har nul på diagonalen, så må der gælde, at $\sum_{i=1}^{n}\lambda_i = 0$. For at bevise at $\lambda_i = 1$ er den største egenværdi, antager vi nu, at et $\lambda_l > 1$. For vektoren $\mathbf{x}$ gælder, at $W\mathbf{x}$ er et vægtet gennemsnit af elementerne i $\mathbf{x}$. Det gør, at

$$(W\mathbf{x})_i \in [x_{\min}, x_{\max}],$$

hvor $i = 1,\dots,n$, og $x_{\min}$ og $x_{\max}$ er henholdsvis den mindste og den største værdi i $\mathbf{x}$. Antagelsen $\lambda_l > 1$ resulterer i, at tilfældet $\lambda_l x_l > x_{\max}$ kan opstå, og dermed holder

ligheden $W\mathbf{x} = \boldsymbol{\lambda}_i\mathbf{x}$ ikke, da $\lambda_l x_l \notin [x_{\min}, x_{\max}]$. Derfor må det nødvendigvis gælde, at $\lambda_l \leq 1$, og dermed at $\lambda_i \leq 1$ for $i = 1,\ldots,n$. Heraf og ved (2.9) følger at $\rho \neq 1$.

Nu ser vi på den mindste egenværdi for W, som vi kalder $\lambda_{\min}$. Vi ved fra (2.10), at der for et k gælder, at $\lambda_k = 1$, så W må derfor have mindst én negativ egenværdi. Vi behøver dog ikke at se nærmere på den mindste egenværdi for W, fordi vi kan sætte $\rho \in (\lambda_{\min}^{-1}, 1)$, idet dette interval som ønsket ikke indeholder nogle af værdierne $\frac{1}{\lambda_i}$ for $i = 1,\ldots,n$.

Vi har nu bevist, at hvis $\rho \in (\lambda_{\min}^{-1}, 1)$, så er $I_n - \rho W$ invertibel. ■

Det er ikke altid tilfældet, at $I_n - \rho W$ er en symmetrisk matrix eller similær med en sådan samt positiv definit, hvilket kan resultere i, at W kan have komplekse egenværdier. Vi undersøger nu, hvornår $I_n - \rho W$ er invertibel og derigennem hvilket interval, der kan vælges for ρ, når der haves komplekse egenværdier.

Lemma 2.13:
Hvis vægtmatricen W har komplekse egenværdier, er $I_n - \rho W$ invertibel, når $\rho \in (\lambda_{\min}^{-1}, 1)$, hvor $\lambda_{\min}^{-1}$ er lig den laveste, negative reelle egenværdi for W.

Bevis:
Vi ser ligesom i beviset til lemma 2.12 på determinanten af $I_n - \rho W$ udtrykt ved

$$|I_n - \rho W| = \prod_{j=1}^{n}(1 - \rho\lambda_j) = \left[\prod_{j=3}^{n}(1 - \rho\lambda_j)\right](1 - \rho\lambda_1)(1 - \rho\lambda_2),$$

hvor λ_i for $i = 1,\ldots,n$ er W's egenværdier. Determinanten er nul, hvis $(1-\rho\lambda_2)(1-\rho\lambda_1) = 0$ for ethvert λ_1 og λ_2, og dermed vil $I_n - \rho W$ ikke være invertibel. Da λ_1 og λ_2 kan være komplekse, skrives de som det konjugerede par

$$\begin{aligned}\lambda_1 &= r + ic\\ \lambda_2 &= r - ic\\ i &= \sqrt{-1},\end{aligned}$$

hvor r og c er reelle tal, og det antages, at $c \neq 0$. De værdier af ρ, der ikke giver invertibel $I_n - \rho W$, kan da findes ved at løse

$$\begin{aligned}0 &= (1 - \rho\lambda_1)(1 - \rho\lambda_2)\\ &= (1 - \rho r - \rho ic)(1 - \rho r + \rho ic)\\ &= 1 - 2\rho r + \rho^2 r^2 - \rho^2 i^2 c^2\\ &= 1 - 2\rho r + \rho^2(r^2 + c^2).\end{aligned} \tag{2.11}$$

Diskriminanten d for (2.11) er

$$\begin{aligned}d &= (-2r)^2 - 4\cdot(r^2 + c^2)\cdot 1\\ &= 4(r^2 - r^2 - c^2)\\ &= -4c^2 < 0.\end{aligned}$$

Matricen $I_n - \rho W$ er ikke invertibel, hvis løsningen er to komplekse værdier for ρ, idet determinanten er negativ. Hvis det på forhånd er antaget, at ρ kun må være reelle tal, vil komplekse egenværdier ikke påvirke, hvorvidt $I_n - \rho W$ er invertibel eller ej.

Hvis W har komplekse egenværdier, sikrer kravet $\rho \in (\lambda_{\min}^{-1}, 1)$, hvor $\lambda_{\min}^{-1}$ er lig den laveste, negative reelle egenværdi for W, at $I_n - \rho W$ er invertibel. ■

Som tidligere kan det ligeledes i tilfældet med komplekse egenværdier antages, at $\rho \in [0,1)$ for at sikre, at $I_n - \rho W$ er invertibel. Det ses, at ved tilfældet med komplekse egenværdier haves der derfor stadig en veldefineret kovariansmatrix for SAR-modellen.

Sætning 2.14:
Kovariansematricen for SAR-modellen givet ved

$$(I_n - \rho W)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T$$

er symmetrisk og positivt definit, hvis $\rho \in (\lambda_{\min}^{-1}, 1)$, hvor $\lambda_{\min}$ er den mindste egenværdi for den rækkestokastiske matrix W, og den er dermed en veldefineret kovariansmatrix.

Bevis:
Først sættes $A = I_n - \rho W$ for overskuelighedens skyld. Matricen $A^T A$ er positivt semidefinit for enhver matrix A, idet

$$\mathbf{x}^T A^T A\mathbf{x} = (A\mathbf{x})^T A\mathbf{x} = \|A\mathbf{x}\|^2 \geq 0,$$

for en $(n \times 1)$-vektor $\mathbf{x}$ jævnfør [Axler, 1997, s. 144]. Vi mangler nu at vise invertibiliteten af $A^T A$. Hvis A er invertibel, så er $A^T A$ invertibel, da

$$\left(A^T A\right)^{-1} = (A)^{-1}\left(A^T\right)^{-1} = (A)^{-1}\left(A^{-1}\right)^T,$$

og dermed er $A^T A$ positivt definit. Jævnfør lemma 2.12 og 2.13 er A invertibel, hvilket gør, at $A^T A$ er invertibel, så

$$A^T A = (I_n - \rho W)^T\left(I_n - \rho W\right) \Rightarrow$$
$$\left(A^T A\right)^{-1} = (I_n - \rho W)^{-1}\left(\left(I_n - \rho W\right)^T\right)^{-1} = (I_n - \rho W)^{-1}\left(\left(I_n - \rho W\right)^{-1}\right)^T$$

er invertibel og dermed positivt definit. Da konstanten $\sigma^2 > 0$, er matricen

$$(I_n - \rho W)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T$$

ligeledes invertibel, og hele kovariansen er således positivt definit.

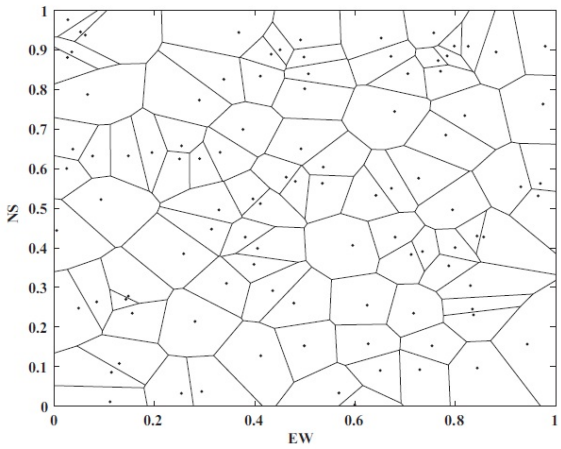
Da der desuden for enhver matrix A gælder, at $A^T A$ og dermed også $\left(A^T A\right)^{-1}$ er en symmetrisk matrix, så er $(I_n - \rho W)^{-1}\sigma^2\left(\left(I_n - \rho W\right)^{-1}\right)^T$ en veldefineret kovariansmatrix. ■

Intervallet for ρ indeholder både negative værdier og værdien nul. Det kan være vanskeligt at fortolke naborelationer, hvis ρ er negativ, så i praksis vælges derfor ofte, at $\rho \in [0,1)$. Værdien $\rho = 0$ vil resultere i, at W udgår fra modellen, som dermed bare er en lineær, normal model, og herved vil der ikke tages højde for naborelationer og ensartethed. Dog medtages nul ofte i intervallet for ρ for at kunne undersøge den mulighed, at der ikke er korrelation mellem observationer.

2.4 Specificering af rumlige vægte

Når man skal specificere sin SAR-model, skal man vælge, hvorledes man vil definere og udregne afstande mellem naboobservationer og således definere naborelationer gennem vægtmatricen W. I stedet for at se på diskrete punkter som f.eks. husstande, kan man tage et område omkring hvert punkt, så hele planen er delt op i forhold til disse. I dette afsnit ser vi på en intuitiv tilgang til det at udforme nabomatricen W baseret på kontinuitet eller nærmeste nabo ved hjælp af euklidiske eller andre metriske afstande. Dog kræver direkte beregning af W $O(n^2)$ operationer, så derfor anvendes ofte andre metoder til at udforme W, der kræver helt ned til $O(n\ln(n))$ operationer. Se appendiks A for yderligere information om tidskompleksitet.

Der er flere måder, hvorpå vægtmatricer kan specificeres. For eksempel kan vi se på et Voronoi-diagram (se figur 2.3), hvor punkter repræsenterer placeringen for en række observationer, og kanterne omkring disse punkter repræsenterer det Voronoiske polygon for punkterne, der angiver grænser for områder, som omgiver de respektive punkter. Disse områder har den egenskab, at alle indre observationer er tættere på det centrale punkt i området end på nogle af de andre punkter i de andre områder. Voronoi-polygoner er altså en metode til at behandle irregulære observationer givet ved punkterne i figur 2.3.



Figur 2.3: Voronoi-diagram med 100 rumlige, stokastiske punkter [LeSage and Pace, 2009, s. 119].

Linjer, der forbinder punkterne mellem de kontinuerte Voronoi-polygoner, angiver kanterne til Delaunay-trekanter. Idet kanterne i Delaunay-trekanterne forbinder punkterne i Voronoi-diagrammet udgør disse trekanter en måde, hvorpå vi kan specificere kontinuitet. Er to punkter forbundet med en kant, anses de for at være naboer. Så givet et sæt af punkter sammen med tilhørende Delaunay-trekanter kan man specificere vægtmatricen W. Et generelt resultat om Voronoi-diagrammer er, at hvert punkt gennemsnitligt har seks naboer, så der vil være tæt på $6n$ Delaunay-trekanter, hvor en trekant udtrykkes ved tre af punkterne. Givet Delaunay-trekanterne kræves kun $O(n)$ operationer til at lave en tyndt besat vægtmatrix W. Vægtmatricen W fundet ud fra Delaunay-trekanter kan anvendes til at finde de m nærmeste naboer for hvert punkt, idet W repræsenterer naboer, W^2 repræsenterer naboer til naboer osv. Naborelationer af forskellige ordener kan

fungere som nabokandidater for de nærmeste naboer. Lad for eksempel

$$Z_4 = W + W^2 + W^3 + W^4.$$

De elementer i Z_4, der ikke er nul, repræsenterer naboer op til den fjerde ordens nabo. Når nabokandidaterne er specificeret, kan vi ud fra koordinater for punkterne udregne afstanden i en givet metrik fra observation i til alle punkterne (observationerne) blandt nabokandidaterne. Sortering af denne korte mængde af distancer mellem nabokandidaterne for at finde de m nærmeste naboer tager ikke lang tid, og størrelsen af nabokandidatsættet er ikke afhængig af n for større n. Hvis vi udfører ovenstående for n observationer, altså n gange i alt, fås en $(n \times m)$-matrix indeholdende information om de m nærmeste naboer for hver observation i. Givet $(n \times m)$-matricen med informationerne kræves kun $O(n)$ operationer for at udregne vægtmatricen W for de m nærmeste naboer. Det vil altså sige, at hvis $(W)_j$ angiver den j'te søjle i vægtmatricen W, så har vi for eksempel, at den første indgang i $(W)_2$ angiver første observations anden nærmeste nabo, og det tiende element i $(W)_5$ er den tiende observations femte nærmeste nabo.

Man kan også definere $(n \times m)$-matricerne $W_{(1)}$, $W_{(2)}$ osv. som heholdsvist den nærmeste nabomatrix, den anden nærmeste nabomatrix osv., så der haves en separat matrix for hver orden af naboerne.

Der findes flere metoder, hvorpå man kan udregne Voronoi-diagrammer og Delaunay-trekanter med en kompleksitet på $O(n\ln(n))$ operationer for data i planen.

2.5 Påvirkninger mellem observationer

En generel regressionsmodel, hvor observationerne er lineære og uafhængige, har formen

$$\mathbf{y} = \sum_{r=1}^{k} \mathbf{x}_r \beta_r + \boldsymbol{\varepsilon}, \tag{2.12}$$

hvor $\mathbf{x}_r$ er den r'te søjle i designmatricen X, mens β_r er den r'te indgang i parametervektoren $\boldsymbol{\beta}$. De partielt afledede for modellen er givet ved

$$\frac{\partial y_i}{\partial x_{ir}} = \frac{\partial}{\partial x_{ir}}\left(x_{ir}\beta_r + \varepsilon_i\right) = \beta_r \qquad \forall \quad i,r$$
$$\frac{\partial y_i}{\partial x_{jr}} = 0 \qquad \forall r \text{ når } j \neq i.$$

Ovenstående kan opfattes som, at informationsmængden for observation nummer i i regression kun består af baggrundsvariable. Modellen (2.12) kan udvides til at omfatte rumlige lags, hvilket vil sige, at informationsmængden udvides med naboobservationernes information.

Sætning 2.15:
De partielt afledede for SAR-modellen er givet ved

$$\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \quad \text{og} \quad \frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij}, \tag{2.13}$$

hvor $S_r(W) = (I_n - \rho W)^{-1}\beta_r$, og β_r er den r'te indgang i parametervektoren $\boldsymbol{\beta}$.

Bevis:
Vi ser på omskrivningen for SAR-modellen

$$\begin{aligned}\mathbf{y} &= \left(I_n - \rho W\right)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= (I_n - \rho W)^{-1} X\boldsymbol{\beta} + (I_n - \rho W)^{-1}\boldsymbol{\varepsilon}.\end{aligned}$$

Det haves, at

$$\begin{aligned}X\boldsymbol{\beta} = \sum_{r=1}^{k} \mathbf{x}_r \beta_r = \sum_{r=1}^{k} \beta_r \mathbf{x}_r \quad \Rightarrow \\ \mathbf{y} = \sum_{r=1}^{k} (S_r(W)\mathbf{x}_r) + V(W)\boldsymbol{\varepsilon},\end{aligned}$$

hvor $\mathbf{x}_r$ er den r'te søjle i designmatricen X, k er antallet af søjler i X, og

$$\begin{aligned}S_r(W) &= V(W)\beta_r \\ V(W) &= (I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \ldots.\end{aligned} \tag{2.14}$$

Her er funktionen $V(W)$ omskrevet efter den geometriske sumformel, da ρ antages at opfylde $|\rho| < 1$. Det ses ved omskrivningen af $V(W)$, at der her inddrages naboer af alle grader i form af potenserne af W i stedet for kun førsteordensnaboer. Hvis den datagenererende proces skrives på matrixform fås, at

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \sum_{r=1}^{k} \begin{bmatrix} S_r(W)_{11} & S_r(W)_{12} & \cdots & S_r(W)_{1n} \\ S_r(W)_{21} & S_r(W)_{22} & \cdots & S_r(W)_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ S_r(W)_{n1} & S_r(W)_{n2} & \cdots & S_r(W)_{nn} \end{bmatrix} \begin{bmatrix} x_{1r} \\ x_{2r} \\ \vdots \\ x_{nr} \end{bmatrix} + V(W)\boldsymbol{\varepsilon}. \tag{2.15}$$

En enkelt række i ligningssystemet kan udtages, så det fås for y_i, at

$$y_i = \sum_{r=1}^{k} (S_r(W)_{i1}x_{1r} + S_r(W)_{i2}x_{2r} + \ldots + S_r(W)_{in}x_{nr}) + V(W)_i\boldsymbol{\varepsilon},$$

således at den partielt afledede kan findes ved

$$\frac{\partial y_i}{\partial x_{ir}} = S_r(W)_{ii} \quad \text{og} \quad \frac{\partial y_i}{\partial x_{jr}} = S_r(W)_{ij}.$$

■

Det ses, at den afledede ikke længere blot er β_r for $i = j$, og at selvom $i \neq j$ er den afledede ikke nødvendigvis nul som i det simple, lineære tilfælde. Denne egenskab fortolkes som at en ændring i baggrundsvariablen for en enkelt observation kan påvirke den afhængige variabel i resten af observationerne. Den afledede i (2.13) beskriver det tilfælde, hvor observation i påvirkes af en ændring i x_{ir}, inklusive de ændringer der stammer fra, at observation i kan påvirke observation j, der så igen påvirker i. Denne form for cirkeleffekt eller feedback kan løbe gennem mere end to observationer, så i for eksempel kan have inflydelse på j, der påvirker k, som så igen påvirker i.

Matricen $S_r(W)$ kan som tidligere nævnt omskrives med den geometriske sumformel, så

$$S_r(W) = (I_n - \rho W)^{-1}\beta_r = (I_n + \rho W + \rho^2 W^2 + \ldots)\beta_r$$

Diagonalen i W^2 er ikke nul som i W, hvilket viser feedback, hvor i er andenordensnabo til sig selv, hvilket vil sige, at i er nabo til sin nabo j, og denne påvirkning vises i $S_r(W)$ via diagonalen i W^2. Graden, af hvor meget i påvirker sig selv gennem sine naboer, afhænger af:

1. Hvordan observationerne er placeret i rummet.
2. Matricen W der afbilder graden af påvirkning mellem observationerne.
3. Parameteren ρ der er et mål for graden af den rumlige afhængighed.
4. Parametervektoren $\boldsymbol{\beta}$.

Diagonalen i matricen $S_r(W)$ kaldes også de direkte påvirkninger, mens resten af elementerne i matricen er de indirekte påvirkninger. For eksempel er det muligt at finde de påvirkninger på **y**, der kommer fra et bestemt område, ved at se på en enkelt søjle i $S_r(W)$, mens man kan se på en enkelt række i $S_r(w)$ for at se, hvilken effekt en ændring har på en enkelt observation. Det er forskelligt hvordan en ændring i en baggrundsvariabel påvirker de forskellige observationer, det vil sige, at y_i påvirkes måske mere end y_j, og derfor kan man i stedet tage gennemsnittet af påvirkningerne for alle observationer. Der er forskellige måder at udregne dette, hvoraf to har samme numeriske værdi, hvilket ses i forbindelse med definition 2.17.

Definition 2.16 (Gennemsnitlig, direkte påvirkning):
Den gennemsnitlige, direkte påvirkning på y_i fra en ændring i x_{ir} er for alle $i = 1,\ldots,n$ givet ved

$$\bar{M}(r)_{direkte} = \frac{tr(S_r(W))}{n}.$$

Den gennemsnitlige, direkte påvirkning på y_i er således middelværdien af diagonalen i $S_r(W)$, da diagonalen er de direkte påvirkninger. Den kan også skrives som

$$\begin{aligned}
\bar{M}(r)_{direkte} &= \frac{tr((I_n - \rho W)^{-1})\beta_r}{n} \Rightarrow \\
\bar{M}_{direkte} &= \begin{bmatrix} \bar{M}(1)_{direkte} \\ \vdots \\ \bar{M}(k)_{direkte} \end{bmatrix} \\
&= \frac{tr((I_n - \rho W)^{-1})\boldsymbol{\beta}}{n} \\
&= \frac{\mathrm{tr}(I_n + \rho W + \rho^2 W^2 + \ldots + \rho^q W^q)\boldsymbol{\beta}}{n} \\
&= n^{-1}\sum_{j=0}^{q} \rho^j \,\mathrm{tr}(W^j)\boldsymbol{\beta} \quad \text{for} \quad q \to \infty
\end{aligned} \tag{2.16}$$

ved brug af (2.14).

Definition 2.17 (Gennemsnitlig, samlede påvirkning):
Den gennemsnitlige, samlede påvirkning er givet ved

$$\bar{M}(r)_{total} = \frac{\iota_n^T S_r(W)\iota_n}{n}.$$

Jævnfør (2.14) kan definition 2.17 også skrives som

$$\bar{M}(r)_{total} = \frac{\iota_n^T (I_n - \rho W)^{-1}\beta_r \iota_n}{n}. \tag{2.17}$$

Den gennemsnitlige, samlede påvirkning kan regnes på to måder, nemlig som

1. Den gennemsnitlige, samlede påvirkning *på* en observation
2. Den gennemsnitlige, samlede påvirkning *fra* en observation

Den gennemsnitlige, samlede påvirkning *på* en observation beskriver, hvordan ændringer i alle observationer påvirker en enkelt observation, og den gennemsnitlige, samlede påvirkning *fra* en observation beskriver, hvordan ændringer i en enkelt observation påvirker alle observationer.

Hvis baggrundsvariablen $\mathbf{x}_r$ i (2.15) ændres til $\mathbf{x}_r + \delta\iota_n$ for alle n observationer, så påvirker det hvert y_i med rækkesummen

$$\delta \sum_{k=1}^{n} S_r(W)_{ik}.$$

De n observationer har derfor en gennemsnitlig, samlet påvirkning på

$$\delta \frac{\iota_n^T S_r(W)\iota_n}{n} = \delta n^{-1} \iota_n^T c_r,$$

hvor $c_r := S_r(W)\iota_n$ er en søjlevektor. Det kaldes den gennemsnitlige, samlede påvirkning *på* en observation. Den gennemsnitlige, samlede påvirkning på en observation kan ses som værende påvirkningen på hver individuelle y_i, der er fremkommet ved en ændring på alle n observationer i den r'te baggrundsvariabel. Der er altså tale om et gennemsnit af rækkesummer.

Hvis baggrundsvariablen i den j'te observation ændres fra x_{jr} til $x_{jr} + \delta$, er den samlede påvirkning på alle y_i givet ved søjlesummen

$$\delta \sum_{k=1}^{n} S_r(W)_{kj}.$$

Den gennemsnitlige påvirkning på hver observation bliver dermed

$$\delta \frac{\iota_n^T S_r(W)\iota_n}{n} = \delta n^{-1} r_r \iota_n,$$

hvor $r_r := \iota_n^T S_r(W)$ er en rækkevektor. Det kaldes den gennemsnitlige, samlede påvirkning *fra* en observation. Den gennemsnitlige, samlede påvirkning fra en observation kan ses som værende påvirkningen på alle y_i'erne, der er fremkommet ved en ændring i den r'te baggrundsvariabel ved den j'te observation y_j. Der er altså tale om et gennemsnit af søjlesummer.

Det ses, at den numeriske værdi af den gennemsnitlige, samlede påvirkning er ens, så den gennemsnitlige, samlede påvirkning er middelværdien af alle $\frac{\partial y_i}{\partial x_{jr}}$ for alle j,i.

Definition 2.18 (Gennemsnitlig, indirekte påvirkning):
Den gennemsnitlige, indirekte påvirkning af y_i fra en ændring i x_{ir} er for alle $i = 1,\ldots,n$ givet ved

$$\bar{M}(r)_{indirekte} = \bar{M}(r)_{total} - \bar{M}(r)_{direkte}.$$

Den gennemsnitllige, samlede påvirkning kan ses som værende gennemsnittet af alle afledede af y_i med hensyn til x_{jr} for alle i og alle j, og den gennemsnitlige, direkte påvirkning kan ses som værende alle egenafledede. Hermed er de indirekte påvirkninger lig med alle afledede fratrukket alle egenafledede.

Det er ikke altid praktisk muligt at udregne $\bar{M}$ ud fra en af de tre formler i definitionerne, da den inverse af $(n \times n)$-matricen $I_n - \rho W$ kan være vanskelig at udregne for store n. Derfor kan man i stedet anvende omskrivningen med den geometriske sumformel og bruge en tilnærmelse af rækken, hvor q er tilstrækkelig høj til at rækken konvergerer tilnærmelsesvist, hvilket vil sige, at

$$(I_n - \rho W)^{-1} \approx I_n + \rho W + \rho^2 W^2 + \ldots + \rho^q W^q.$$

Påvirkningerne ses i $S_r(W)$, der kan udvides til en linearkombination af W, og i praksis forenkles funktionen ud fra samme princip som ovenfor, så

$$S_r(W) \approx (I_n + \rho W + \rho^2 W^2 + \ldots + \rho^q W^q)\beta_r.$$

En ændring i en baggrundsvariabel påvirker observationer tæt på ændringen mere end observationer længere væk, det vil sige naboer af højere orden påvirkes mindre des større ordenen er.

Sætning 2.19:
Den gennemsnitlige, samlede påvirkning for SAR-modellen er givet ved

$$\bar{M}(r)_{total} = \left(1 - \rho\right)^{-1} \beta_r,$$

hvor β_r er den r'te ingang i parametervektoren $\boldsymbol{\beta}$.

Bevis:
Ved anvendelse af definition 2.17 fås det, at

$$\bar{M}(r)_{total} = n^{-1}\iota_n^T S_r(W)\iota_n = n^{-1}\iota_n^T (I_n - \rho W)^{-1}\beta_r \iota_n.$$

Den geometriske sumformel kan nu anvendes, så

$$\begin{aligned} n^{-1}\iota_n^T(I_n-\rho W)^{-1}\beta_r\iota_n &= n^{-1}\iota_n^T\left(\sum_{i=0}^{\infty}\rho^j W^j\right)\beta_r\iota_n \\ &= n^{-1}\left(\sum_{i=0}^{\infty}\rho^j\iota_n^T W^j\iota_n\right)\beta_r \\ &= n^{-1}\left(\sum_{i=0}^{\infty}\rho^j n\right)\beta_r, \end{aligned}$$

da vægtmatricen W er rækkestokastisk. Anvendes den geometriske sumformel igen, fås

$$n^{-1}\left(\sum_{i=0}^{\infty}\rho^j n\right)\beta_r = \left(1-\rho\right)^{-1}\beta_r.$$

■

Maximum Likelihood-estimation for SAR-modellen 3

Dette kapitel er skrevet på baggrund af [LeSage and Pace, 2009, kapitel 3], [Olofsson, 2005, afsnit 6.4.2] og [Azzalini, 1996, s. 22-29].

Parametrene i en model for et datasæt kan estimeres ved brug af mindste kvadraters metode, men hvis der haves rumligt laggede, afhængige variable, vil maximum likelihood-estimation give rumlige parametre og standardfejl, hvor afvigelsen mellem model og data er mindre end ved mindste kvadraters metode.

3.1 Maximum Likelihood-estimation

Når man ønsker at estimere parametrene i en model, anvendes ofte maximum likelihood-estimation. Der ses da på en sandsynlighedsfordeling udtrykt ved en model for de pågældende data, og denne fordeling ønskes maksimeret. Når vi vil maksimere likelihoodfunktionen i praksis, kan det ofte være en fordel at se på loglikelihoodfunktionen i stedet. Det kan lade sig gøre, idet ln er en stigende funktion, hvilket vil sige, at funktionen bevarer maksimum fra likelihoodfunktionen, når den anvendes på denne. Fordelen ved at se på loglikelihoodfuntionen er, at den laver produkter, som likelihoodfunktioner ofte består af jævnfør [Olofsson, 2005, s. 329], om til summer, som er meget lettere at håndtere, når der skal maksimeres, idet der ofte maksimeres ved at differentiere og sætte lig med nul.

Definition 3.1 (Likelihoodfunktionen):
Likelihoodfunktionen $l(\boldsymbol{\theta})$ er givet ved

$$l = l(\boldsymbol{\theta}|\mathbf{y}) = f_{\boldsymbol{\theta}}(\mathbf{y}),$$

hvor $\boldsymbol{\theta} = \theta_1, \theta_2, \ldots$ er parametre, og $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \ldots & y_n \end{bmatrix}^T$ er en datavektor. Funktionen $f_{\boldsymbol{\theta}}(\mathbf{y})$ er en tæthedsfunktion fra en fordeling med pmf eller pdf $f_{\boldsymbol{\theta}}$. Den vektor $\hat{\boldsymbol{\theta}}$, der giver den maksimale værdi for likelihoodfunktionen, er det maksimale likelihood-estimat, MLE.

I likelihoodfunktionen $l(\boldsymbol{\theta}|\mathbf{y})$ anses data for at være kendt, mens modelparametrene er de ukendte, og i tætheden $f_{\boldsymbol{\theta}}(\mathbf{y})$ er data ukendt, mens parametrene kendes, således at ny data kan simuleres fra denne. For SAR-modellen er $\boldsymbol{\theta} = (\boldsymbol{\beta}, \rho, \sigma^2)$, og hvis $\alpha\iota_n$ er trukket ud af designmatricen X, så vil α også indgå i $\boldsymbol{\theta}$.

Når likelihoodfunktionen er kendt, kan loglikelihoodfunktionen findes.

Definition 3.2 (Loglikelihoodfunktionen):
Loglikelihoodfunktionen er givet ved

$$L = L(\boldsymbol{\theta}|\mathbf{y}) = \ln l(\boldsymbol{\theta}|\mathbf{y}) = \ln(f_{\boldsymbol{\theta}}(\mathbf{y})),$$

hvor $l(\boldsymbol{\theta})$ er en likelihoodfunktion.

Ved maximum likelihood-estimation ses der på, hvornår sandsynligheden for de forskellige parametre i en model er maksimal, hvilket gøres ved at se på loglikelihoodfunktionen for parametrene. Loglikelihoodfunktionen differentieres partielt med hensyn til alle de forskellige parametre i den pågældende model, der ønskes estimeret, og alle disse førsteordensafledede sættes lig nul for at finde maksimum for hver af de pågældende parametre. Idet det er kompliceret at estimere alle parametre simultant, ses der på én parameter ad gangen, mens resten af parametrene holdes fast eller ses som funktioner af den variable parameter. En loglikelihoodfunktion med hensyn til kun én af parametrene i en model, således at de andre parametre holdes fast, kaldes en profilloglikelihoodfunktion.

Definition 3.3 (Profilloglikelihoodfunktionen):
En profilloglikelihood er loglikelihoodfunktionen med hensyn til kun én parameter og er givet ved

$$L_p(\theta_j) = \ln(f_{\theta_j}(y_i)),$$

hvor θ_j er givet for ét $j = 1,\ldots,$ således at θ_j ses som den eneste variable parameter, mens alle andre parametre holdes fast.

Når der arbejdes med SAR-modellen, ses der ofte især på profilloglikelihoodfunktionen med hensyn til ρ, det vil sige $L_p(\rho)$, således at resten af parametrene $\boldsymbol{\beta}$ og σ^2 holdes fast, idet ρ er den parameter, der angiver rumlig afhængighed.

Fremgangsmåden ved maximum likelihood-estimation for en model med flere ukendte parametre er for hver ukendt parameter, at

1. finde $l(\boldsymbol{\theta})$ og dernæst $L(\boldsymbol{\theta}) = \log l(\boldsymbol{\theta})$
2. finde $L_p(\theta_i)$
3. differentiere $L_p(\theta_i)$ med hensyn til θ_i og sæt lig nul
4. finde løsning givet ved $\theta_i = \hat{\theta}_i$
5. Undersøge hvorvidt $L''(\hat{\theta}_i) < 0$ for at sikre et maksimum.

3.2 Parameterestimering for SAR-modellen

Vi vil nu se på, hvordan parametrene i en SAR-model estimeres ved hjælp af først mindste kvadraters metode og derefter maximum likelihood-estimation. Vi har brug for et lemma til at bevise estimatets form for $\boldsymbol{\beta}$.

Lemma 3.4:
Hvis X er en $(n \times k)$-designmatrix med fuld søjlerang, så er $(k \times k)$-matricen $X^T X$ positivt definit, og matricen $(X^T X)^{-1}$ eksisterer.

Bevis:
Idet X har fuld søjlerang, og alle søjler i X dermed er uafhængige, gælder der for alle $\mathbf{v} \neq \mathbf{0} \in \mathbb{R}^k$, at

$$\begin{aligned}\mathbf{v}^T \left(X^T X\right)\mathbf{v} &= \left(\mathbf{v}^T X^T\right)(X\mathbf{v}) \\ &= (X\mathbf{v})^T (X\mathbf{v}) \\ &= \|X\mathbf{v}\|^2 > 0,\end{aligned}$$

og dette betyder ifølge [Axler, 1997, s. 144], at matricen $X^T X$ er positivt definit. Dermed gælder der, at matricen $(X^T X)^{-1}$ eksisterer. ■

Sætning 3.5:
Ved mindste kvadraters metode er estimaterne for parametervektoren $\boldsymbol{\beta}$, variansen σ^2 og kovariansmatricen Ω i SAR-modellen givet ved

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (X^T X)^{-1} X^T \left(I_n - \rho^* W\right)\mathbf{y} \\ \hat{\sigma}^2 &= \frac{\left\|e(\rho^*)\right\|}{n} = n^{-1} e(\rho^*)^T e(\rho^*) \\ \hat{\Omega} &= (I_n - \rho^* W)^{-1} \hat{\sigma}^2 \left((I_n - \rho^* W)^{-1}\right)^T,\end{aligned}$$

hvor ρ^* er den ægte værdi for ρ, og

$$e(\rho^*) = \mathbf{y} - \rho^* W\mathbf{y} - X\hat{\boldsymbol{\beta}}.$$

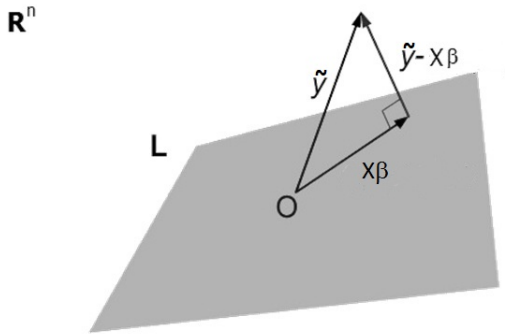
Bevis:
Det antages, at den ægte værdi for ρ kendes som ρ^*, så SAR-modellen er givet ved

$$\begin{aligned}\mathbf{y} &= \rho^* W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \Rightarrow \\ (I_n - \rho^* W)\mathbf{y} &= X\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ \boldsymbol{\varepsilon} &\sim N_n(\mathbf{0}, \sigma^2 I_n).\end{aligned} \tag{3.1}$$

Vi sætter nu $\tilde{\mathbf{y}} := (I_n - \rho^* W)\mathbf{y}$, så $\tilde{\mathbf{y}} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Middelværdivektoren for observationerne $\tilde{\mathbf{y}}$ er givet ved $X\boldsymbol{\beta}$, hvor vektoren $\boldsymbol{\beta} \in \mathbb{R}^k$, og $X\boldsymbol{\beta}$ tilhører et k-dimensionalt underrum $L \subset \mathbb{R}$, der er udspændt af søjlerne i $(n \times k)$-designmatricen X.

Det antages, at vi har en mængde observationer $\tilde{\mathbf{y}} = \begin{bmatrix}\tilde{y}_1 & \tilde{y}_2 & \dots & \tilde{y}_n\end{bmatrix}^T$ og skal bestemme parametrene i en passende model. Når en specifik model er valgt ønskes en bestemmelse

af $\boldsymbol{\beta}$, der beskriver dataene $\tilde{\mathbf{y}} = \begin{bmatrix} \tilde{y}_1 & \tilde{y}_2 & \dots & \tilde{y}_n \end{bmatrix}^T$ bedst muligt. Vi må nødvendigvis have, at middelværdien $X\boldsymbol{\beta}$ fremkommer ved at projektere $\tilde{\mathbf{y}}$ ind på underrummet L, og dermed gælder det, at $(\tilde{\mathbf{y}} - X\boldsymbol{\beta}) \in L^{\perp}$. Dette er illustreret på figur 3.1, og en sådan projektion svarer til, at vi anvender mindste kvadraters metode.



Figur 3.1: Projektion af data $\mathbf{y}$ på underrummet L.

Da L er udspændt af søjlerne i matricen X, kan enhver vektor i L skrives på formen $X\mathbf{v}$ for et $\mathbf{v} \in \mathbb{R}^k$. Eftersom $(\tilde{\mathbf{y}} - X\boldsymbol{\beta}) \in L^{\perp}$, medfører dette, at

$$\begin{aligned} 0 &= (X\mathbf{v})^T(\tilde{\mathbf{y}} - X\boldsymbol{\beta}) \\ &= \mathbf{v}^T X^T(\tilde{\mathbf{y}} - X\boldsymbol{\beta}) \\ &= \mathbf{v}^T(X^T\tilde{\mathbf{y}} - X^T X\boldsymbol{\beta}) \end{aligned}$$

for alle $\mathbf{v} \in \mathbb{R}^k$. Idet vi ikke ønsker den trivielle løsning, hvor $\mathbf{v} = 0$, haves der, at

$$\begin{aligned} 0 &= \mathbf{v}^T(X^T\tilde{\mathbf{y}} - X^T X\boldsymbol{\beta}) \Rightarrow \\ 0 &= X^T\tilde{\mathbf{y}} - X^T X\boldsymbol{\beta} \\ X^T X\boldsymbol{\beta} &= X^T\tilde{\mathbf{y}} \Rightarrow \\ \boldsymbol{\beta} &= (X^T X)^{-1} X^T\tilde{\mathbf{y}}, \end{aligned}$$

idet vi fra lemma 3.4 ved, at $X^T X$ er invertibel. Idet $\tilde{\mathbf{y}} := \left(I_n - \rho^* W\right)^{-1}\mathbf{y}$, fås estimatet for $\boldsymbol{\beta}$ til

$$\hat{\boldsymbol{\beta}} = (X^T X)^{-1} X^T\left(I_n - \rho^* W\right)\mathbf{y}.$$

Fejlvektoren for anvendelse af ovenstående med hensyn til ρ^* er givet ved dataene fratrukket værdierne fra den estimerede model for dataene, så

$$e(\rho^*) = \mathbf{y} - \rho^* W\mathbf{y} - X\hat{\boldsymbol{\beta}},$$

Divideres fejlvektoren med antallet af observationer n, fås estimatet for fejlleddets varians for SAR-modellen til

$$\hat{\sigma}^2 = \frac{\left\| e(\rho^*) \right\|}{n} = n^{-1} e(\rho^*)^T e(\rho^*),$$

og estimatet for kovariansmatricen er derfor givet ved

$$\hat{\Omega} = (I_n - \rho^* W)^{-1} \hat{\sigma}^2 \left((I_n - \rho^* W)^{-1}\right)^T. \qquad \blacksquare$$

Hvis $\alpha \iota_n$ indgår i (3.1), kan matricen $Z := \begin{bmatrix} \iota_n & X \end{bmatrix}$ og vektoren $\boldsymbol{\delta} := \begin{bmatrix} \alpha & \boldsymbol{\beta} \end{bmatrix}^T$ defineres, og dermed kan parameteren $\boldsymbol{\delta}$ estimeres som

$$\hat{\boldsymbol{\delta}} = (Z^T Z)^{-1} Z^T (I_n - \rho^* W)\mathbf{y}.$$

Vi vil nu se på maximum likelihood-estimation. Omskrivningerne, som er vist herover, viser, at maximum likelihood-estimationen kan reduceres til maksimering kun med hensyn til parameteren ρ, idet $\hat{\boldsymbol{\beta}}$ kun afhænger af ρ. Dette gøres ved at anvende profilloglikelihoodfunktionen $L_p(\rho)$. Profilloglikelihoodfunktionen $L_p(\rho)$ er konstrueret med hensyn ρ, således at parametrene $\boldsymbol{\beta}$ og σ^2 holdes fast ellers ses som funktioner ρ.

Hvis problemet skulle løses for den fulde loglikelihoodfunktion for SAR-modellen, skulle de førsteordensafledede med hensyn til $\boldsymbol{\beta}$, σ^2 og ρ sættes lig nul og løses simultant for alle tre parametre. Men som ovenstående viser, er en anden metode at bruge profilloglikelihoodfunktionen $L_p(\rho)$. Profilloglikelihoodfunktionerne for $\boldsymbol{\beta}$ og σ^2 skrives henholdsvis som $L_p(\boldsymbol{\beta}(\rho))$ og $L_p(\sigma^2(\rho))$, så de hver er afhængig af datasættet og den ukendte parameter ρ. Således fås for SAR-modellen, at den eneste parameter, der skal estimeres i profilloglikelihoodfunktionen, er ρ, og den fundne værdi $\hat{\rho}$ kan derefter indsættes i de to andre funktioner, så vi har udtrykkene $L_p(\boldsymbol{\beta}(\hat{\rho}))$ og $L_p(\sigma^2(\hat{\rho}))$, der giver de resterende maximum likelihood-estimater. Det skal bemærkes, at estimering af parametrene giver samme resultat, om der ses på den fulde loglikelihood eller på profilloglikelihoodfunktionen, men ved profilloglikelihoodfunktionen får vi reduceret et multivariat problem til et univariat problem.

Sætning 3.6:
For SAR-modellen er likelihoodfunktionen l og loglikelihoodfunktionen $L = \ln l$ givet ved henholdsvis

$$l = (2\pi\sigma^2)^{-\frac{n}{2}} \cdot \left|I_n - \rho W\right| \cdot \exp\left\{ -\frac{\left(\left(I_n - \rho W\right)\mathbf{y} - X\boldsymbol{\beta}\right)^T \left((I_n - \rho W)\mathbf{y} - X\boldsymbol{\beta}\right)}{2\sigma^2} \right\}$$

og

$$L = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln\left|I_n - \rho W\right| - \frac{e^T e}{2\sigma^2}, \tag{3.2}$$

hvor $e = \mathbf{y} - \rho W\mathbf{y} - X\boldsymbol{\beta}$.

Bevis:
SAR-modellen er ifølge sætning 2.8 normalfordelt med

$$N_n\left((I_n - \rho W)^{-1} X\boldsymbol{\beta}, (I_n - \rho W)^{-1}\sigma^2 \left(\left(I_n - \rho W\right)^{-1}\right)^T\right).$$

Ifølge appendiks B.1 har SAR-modellen dermed tætheden

$$p(\mathbf{y}) = \left(\sqrt{(2\pi)^n \left|\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|}\right)^{-1} \tag{3.3}$$
$$\cdot \exp\left\{-\frac{1}{2}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right)^{-1}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\right\}.$$

Udregnes den første faktor heri, fås

$$\begin{aligned}\left(\sqrt{(2\pi)^n \left|\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|}\right)^{-1} &= \left((2\pi)^n\left|\left(I_n-\rho W\right)^{-1}\right|\left|\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|\right)^{-\frac{1}{2}}\\ &= (2\pi)^{-\frac{n}{2}}\left|I_n-\rho W\right|^{\frac{1}{2}}\left(\left(\sigma^2\right)^n\left|\left(\left(I_n-\rho W\right)^{-1}\right)^T\right|\right)^{-\frac{1}{2}}\\ &= (2\pi)^{-\frac{n}{2}}\left|I_n-\rho W\right|^{\frac{1}{2}}\sigma^{-n}\left|I_n-\rho W\right|^{\frac{1}{2}}\\ &= (2\pi)^{-\frac{n}{2}}\left|I_n-\rho W\right|\sigma^{-n}\\ &= \left(2\pi\sigma^2\right)^{-\frac{n}{2}}\left|I_n-\rho W\right|,\end{aligned}$$

idet $\det A = \det A^T$ og $\det(rA) = r^n \det A$ for en $(n\times n)$-matrix A og en konstant r.

Indholdet i eksponenten fra (3.3) giver

$$\begin{aligned}&-\frac{1}{2}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)^{-1}\sigma^2\left(\left(I_n-\rho W\right)^{-1}\right)^T\right)^{-1}\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\\ &\quad= -\frac{1}{2}\left(\mathbf{y}-(I_n-\rho W)^{-1}X\boldsymbol{\beta}\right)^T\left(I_n-\rho W\right)^T\left(\sigma^2\right)^{-1}\left(I_n-\rho W\right)\left(\mathbf{y}-\left(I_n-\rho W\right)^{-1}X\boldsymbol{\beta}\right)\\ &\quad= -\frac{1}{2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\sigma^{-2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\\ &\quad= -\frac{\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)}{2\sigma^2}.\end{aligned}$$

Hermed er likelihoodfunktionen for SAR-modellen givet ved

$$l = (2\pi\sigma^2)^{-\frac{n}{2}}\cdot\left|I_n-\rho W\right|\cdot\exp\left\{-\frac{\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left((I_n-\rho W)\mathbf{y}-X\boldsymbol{\beta}\right)}{2\sigma^2}\right\}.$$

Loglikelihoodfunktionen $L = \ln l$ er da givet ved

$$L = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln\left|I_n-\rho W\right| - \frac{e^T e}{2\sigma^2}$$

med

$$\begin{aligned}e &= \mathbf{y}-\rho W\mathbf{y}-X\boldsymbol{\beta}\\ \rho &\in (\lambda_{\min}^{-1},1),\end{aligned}$$

hvor $\lambda_{\min}^{-1}$ er den mindste egenværdi for vægtmatricen W. ■

En normalfordeling har kun en veldefineret likelihoodfunktion, der gør det ud for tætheden, hvis varians-kovarians-matricen er positiv definit og symmetrisk, så derfor er betingelsen $\rho \in (\lambda_{\min}^{-1},1)$ nødvendig, idet vi ønsker en tæthedsfunktion for de data, som SAR-modellen beskriver. I praksis anvendes som regel, at $\rho \in [0,1)$, idet der antages, at dataene er positivt korrelerede eller helt ukorrelerede. Profilloglikelihoodfunktionen med hensyn til ρ kan findes ved at omskrive (3.2), så parametrene $\boldsymbol{\beta}$ og σ^2 holdes fast, så

$$L_p(\rho) = \kappa + \ln\left|I_n - \rho W\right| - \frac{n}{2} S(\rho), \tag{3.4}$$

hvor

$$\begin{aligned}
S(\rho) &= e(\rho)^T e(\rho) = e_0^T e_0 - 2\rho e_0^T e_d + \rho^2 e_d^T e_d \\
e(\rho) &= e_0 - \rho e_d \\
e_0 &= \mathbf{y} - X\boldsymbol{\beta}_0 \\
e_d &= W\mathbf{y} - X\boldsymbol{\beta}_d \\
\boldsymbol{\beta}_0 &= (X^T X)^{-1} X^T \mathbf{y} \\
\boldsymbol{\beta}_d &= (X^T X)^{-1} X^T W\mathbf{y},
\end{aligned}$$

og $\kappa = -\frac{n}{2}\ln(2\pi\sigma^2)$, som er en konstant uafhængig af ρ, og $\left|I_n - \rho W\right|$ er determinanten af matricen $I_n - \rho W$. Derudover er $e(\rho)$ en fejlvektor, der er afhængig af ρ, ligesom $L_p(\rho)$ er afhængig af ρ.

Det at maksimere profilloglikelihoodfunktionen med hensyn til ρ kan gøres enklere ved brug af en $(q \times 1)$-vektor af værdier fra intervallet $[\rho_{\min},\rho_{\max}]$, så $\boldsymbol{\rho} = \begin{bmatrix}\rho_1 & \dots & \rho_q\end{bmatrix}^T$, der er mulige værdier for ρ. Det giver, at

$$\begin{bmatrix} L_p(\rho_1) \\ L_p(\rho_2) \\ \vdots \\ L_p(\rho_q) \end{bmatrix} = \kappa\iota_n + \begin{bmatrix} \ln\left|I_n - \rho_1 W\right| \\ \ln\left|I_n - \rho_2 W\right| \\ \vdots \\ \ln\left|I_n - \rho_q W\right| \end{bmatrix} - \frac{n}{2} \begin{bmatrix} S(\rho_1) \\ S(\rho_2) \\ \vdots \\ S(\rho_q) \end{bmatrix}. \tag{3.5}$$

Her udregnes værdierne $e_0^T e_0$, $e_d^T e_d$ og $(k \times 1)$-vektorerne $\boldsymbol{\beta}_0$ og $\boldsymbol{\beta}_d$ først, så der kun mangler værdien for ρ for at udregne $e(\rho)$ og dermed $S(\rho)$. Estimatet $\hat{\rho}$ findes som den optimale værdi af ρ, givet ved ρ_i, der er den værdi fra følgen $\rho_1,\dots,\rho_q$, som maksimerer (3.5). Når den kendes, kan estimaterne for $\boldsymbol{\beta},\sigma^2$ og varians-kovarians-matricen for fejlleddet $\hat{\Omega}$ udregnes som

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= \boldsymbol{\beta}_0 - \hat{\rho}\boldsymbol{\beta}_d \\
\hat{\sigma}^2 &= n^{-1} S(\hat{\rho}) \\
\hat{\Omega} &= (I_n - \hat{\rho}W)^{-1} \hat{\sigma}^2 \left((I_n - \hat{\rho}W)^{-1}\right)^T,
\end{aligned}$$

hvor sidste lighed ses ved at sætte estimaterne for σ^2 og ρ ind i kovariansmatricen i sætning 2.8.

Vi har derfor at gøre med en loglikelihoodfunktion, der kombinerer en kvadratsum af fejlled, $S(\rho)$, med et logdeterminant-led, $\left|I_n - \rho W\right|$, hvor sidstnævnte fungerer som

en form for straffunktion, der sørger for, at maximum likelihood-estimatet for ρ ikke bliver lig med et estimat, der kun er baseret på at minimere den transformerede sum af kvadrerede fejl $S(\rho)$. Desuden giver inddragelsen af udvalgte værdier vektoren $\boldsymbol{\rho} = \begin{bmatrix}\rho_1 & \dots & \rho_q\end{bmatrix}^T$ sikkerheden for et globalt og ikke kun lokalt maksimum, da hele intervallet, hvor ρ kan forekomme, dækkes.

Udover at anvende den metode, hvor ρ ses som en vektor, kan Newtons metode med numeriske afledede anvendes. Denne metode kan være en fordel at anvende, idet vi både får maximum likelihood-estimatet for ρ og den andenordensafledede af profilloglikelihoodfunktionen ved det estimerede $\hat{\rho}$, så et maksimum kan sikres. Det numeriske estimat for den andenordensafledede giver nemlig sammen med anden information en måde, hvorpå vi kan finde et numerisk estimat for varians-kovariansmatricen for en models parametre, hvilket vi vil se på senere i afsnit 5.4.

Det er nu blevet vist, hvordan det er muligt at finde estimater for parameteren ρ ved brug af bivariat og univariat maksimering af profilloglikelihoodfunktionen $L_p(\rho)$ med fast $\boldsymbol{\beta}$ og σ^2, hvor maximum likelihood-estimater for parametrene $\boldsymbol{\beta}$ og σ^2 kan findes ved at anvende maximum likelihood-estimatet for $\hat{\rho}$.

Modeller med to vægtmatricer

SAR-modellen kan indeholde mere end én vægtmatrix.

Definition 3.7 (SAR-model med to vægtmatricer):
SAR-modellen med to vægtmatricer har formen

$$\begin{aligned}\mathbf{y} &= \rho_1 W_1\mathbf{y} + \rho_2 W_2\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}\\ \boldsymbol{\varepsilon} &\sim N_n(\mathbf{0},\sigma^2 I_n),\end{aligned}$$

hvor den første vægtmatrix W_1 afbilder påvirkninger fra naboobservationer i et afgrænset område, og W_2 afbilder påvirkninger fra naboobservationer i et andet afgrænset område.

For eksempel kan W_1 afbilde påvirkninger på området $\mathbf{y}$ fra byer i samme kommune som $\mathbf{y}$, mens W_2 er påvirkningerne fra byer i nabokommunerne.

Således tillades seperate påvirkninger af designmatricen X med hensyn til nabobyer inden for samme kommune og med hensyn til påvirkningerne fra nabobyer i nabokommuner. Nu haves dog en situation, hvor der skal laves maximum likelihood-estimation over intervallet for passende værdier for to variable, nemlig ρ_1 og ρ_2. Hvis der ønskes en maximum likelihood-estimation af profilloglikelihoodfunktionen med hensyn til ρ_1 og ρ_2, involverer dette en udregning eller estimation af logdeterminantleddet givet ved $\left|I_n - \rho_1 W_1 - \rho_2 W_2\right|$ over et bivariat gitter med passende værdier for både ρ_1 og ρ_2. De passende værdier er indeholdt i en matrix og ikke en vektor som ved kun et ρ.

Monte Carlo-approksimation 4

Dette kapitel er skrevet på baggrund af [LeSage and Pace, 2009, kapitel 4] og diverse velkendte regneregler.

4.1 Monte Carlo-approksimation af logdeterminanten

Monte Carlo-approksimation anvendes i forbindelse med approksimation af middelværdier og udføres ved hjælp af gennemsnittet for et antal simulationer. Monte Carlo-approksimation tages i brug, når vi kan udtrykke karakteristika for de givne data ved hjælp af en middelværdi for disse. Vi kan i forbindelse med rumligt afhængige modeller anvende Monte Carlo-approksimation til at finde en approksimation af logdeterminanten. Metoden kræver hverken lang beregningstid eller stor computerhukommelse, og desuden resulterer metoden kun i en lille fejl i estimatet for ρ.

Logdeterminanten af matricen $I_n - \rho W$ kan udtrykkes som sporet af logaritmen til matricen, men først vises to lemmaer.

Lemma 4.1:
For en eksponentialmatrix e^A gælder det, at

$$\left|e^A\right| = e^{\mathrm{tr}(A)},$$

hvor A er en $(n \times n)$-matrix.

Bevis:
Ved hjælp af jordan kanonisk form fås, at

$$J = V^{-1}AV \Rightarrow A = VJV^{-1}.$$

Matricen J er enten en diagonalmatrix eller en diagonalmatrix med en superdiagonal, men da determinanten for en diagonalmatrix og en øvre triangulær matrix begge er givet ved produktet af diagonalelementerne, nøjes vi med at se på tilfældet, hvor A er diagonaliserbar, så

$$A = VJV^{-1} = V\Lambda_A V^{-1},$$

hvor Λ_A er en diagonalmatrix med A's egenværdier på diagonalen, og V består af A's ortonormale egenvektorer. Så haves det per definition af e^A, at

$$\begin{aligned} e^A &= \sum_{k=0}^{\infty} \frac{A^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{(V\Lambda_A V^{-1})^k}{k!} \\ &= \sum_{k=0}^{\infty} \frac{V\Lambda_A^k V^{-1}}{k!} \\ &= V\left(\sum_{k=0}^{\infty} \frac{\Lambda_A^k}{k!}\right) V^{-1} \\ &= Ve^{\Lambda_A}V^{-1} \Rightarrow \\ \left|e^A\right| &= \left|Ve^{\Lambda_A}V^{-1}\right| = |V|\left|e^{\Lambda_A}\right|\left|V^{-1}\right| = \left|e^{\Lambda_A}\right|. \end{aligned} \tag{4.1}$$

Idet Λ_A har A's egenværdier på diagonalen fås, at

$$\left|e^{\Lambda_A}\right| = e^{\lambda_1} \cdot e^{\lambda_2} \cdots e^{\lambda_n} = e^{\lambda_1+\lambda_2+\ldots+\lambda_n} = e^{\mathrm{tr}(\Lambda_A)}. \tag{4.2}$$

Ved at sætte resultaterne (4.1) og (4.2) sammen, fås

$$\left|e^A\right| = e^{\mathrm{tr}(A)},$$

idet $\mathrm{tr}(A) = \sum_{i=1}^{n} \lambda_i$. ■

Lemma 4.2:
For logdeterminanten af matricen $I_n - \rho W$ gælder der, at

$$\ln|I_n - \rho W| = -\sum_{j=1}^{\infty} \frac{\rho^j \,\mathrm{tr}(W^j)}{j},$$

når W er rækkestokastisk, og $|\rho| < 1$.

Bevis:
Fra lemma 4.1 haves, at der for en eksponentialmatrix e^A gælder, at

$$\left|e^A\right| = e^{\mathrm{tr}(A)} \Rightarrow \mathrm{tr}(A) = \ln\left|e^A\right|.$$

Hvis vi sætter $A = \ln(I_n - \rho W)$, fås det, at

$$\mathrm{tr}\left(\ln\left(I_n - \rho W\right)\right) = \ln\left|e^{\ln(I_n-\rho W)}\right| = \ln\left|I_n - \rho W\right|. \tag{4.3}$$

Der tages nu udgangspunkt i en Taylorudvikling for matricer, idet der ses på Taylorudviklingen for udtrykket $\ln(I_n - B)$, for hvilket der gælder, at

$$\ln(I_n - B) = -\sum_{j=1}^{\infty} \frac{B^j}{j},$$

hvor ln kræver, at der skal gælde, at matricen $I_n - B$ er positivt definit. Idet vi sætter $B = \rho W$, fås det, at

$$\ln(I_n - \rho W) = -\sum_{j=1}^{\infty} \frac{(\rho W)^j}{j} = -\sum_{j=1}^{\infty} \frac{\rho^j W^j}{j}. \tag{4.4}$$

Hvis (4.3) og (4.4) sammenholdes, haves det, at

$$\ln\left|I_n - \rho W\right| = \operatorname{tr}\left(\ln\left(I_n - \rho W\right)\right) = \operatorname{tr}\left(-\sum_{j=1}^{\infty} \frac{\rho^j W^j}{j}\right) = -\sum_{j=1}^{\infty} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}.$$

■

Vi kan nu bevise, at logdeterminanten af matricen $I_n - \rho W$ kan udtrykkes som sporet af logaritmen til matricen.

Sætning 4.3:
Logdeterminanten $\ln\left|I_n - \rho W\right|$ kan approksimeres ved

$$\ln\left|I_n - \rho W\right| \approx -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}, \tag{4.5}$$

når W er rækkestokastisk, og $|\rho| < 1$.

Bevis:
Fra lemma 4.1 haves det, at

$$\left|e^A\right| = e^{\operatorname{tr}(A)} \Rightarrow \ln\left|e^A\right| = \operatorname{tr}(A) \Rightarrow \ln|B| = \operatorname{tr}(\ln(B)),$$

hvor $B = e^A$. Derfor haves det, at

$$\ln\left|I_n - \rho W\right| = \operatorname{tr}\left(\ln\left(I_n - \rho W\right)\right).$$

Vi ved desuden fra lemma 4.2, at logaritmen til determinanten af matricen $I_n - \rho W$ kan skrives som

$$\ln\left|I_n - \rho W\right| = -\sum_{j=1}^{\infty} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}, \tag{4.6}$$

således at logdeterminanten er en vægtet række af spor af potenser af W. Logaritmen er også defineret over det komplekse plan med undtagelse af ln(0), så (4.6) er ligeledes defineret for kompleks ρ og for ikke-symmetriske matricer W, der kan have komplekse egenværdier.

Man kan opdele den uendelige række i (4.6) i en endelig række af lavere orden og et restled R, der består af højereordens, uendelige udtryk, så

$$\ln\left|I_n - \rho W\right| = -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j} - \sum_{j=m+1}^{\infty} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j} = -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j} - R,$$

hvor m er en fastsat øvre grænse for approksimationen.

Det ses, at hvis R er lille, så kan logdeterminanten approksimeres ved en endelig række af lavere orden, så det fås, at

$$\ln\left|I_n - \rho W\right| \approx -\sum_{j=1}^{m} \frac{\rho^j \operatorname{tr}\left(W^j\right)}{j}.$$

■

Det kan tage lang tid først at udregne W^j for derefter at beregne sporet af W^j direkte, da sådanne udregninger kan kræve op til $O(n^3)$ operationer, når W er tæt besat. Der findes dog andre metoder til at approksimere sporet af W^j. Vi ser nu på et simpelt tilfælde i to dimensioner.

En fordel ved Monte Carlo-approksimation er, at vi nemt kan opdatere udregningerne for enhver værdi af ρ. Vi kan for eksempel sætte $\mathbf{a} = \begin{bmatrix} -\frac{\operatorname{tr}(W)^{(1)}}{1} & -\frac{\operatorname{tr}(W)^{(2)}}{2} & \dots & -\frac{\operatorname{tr}(W)^{(m)}}{m} \end{bmatrix}$ og $\mathbf{b} = \begin{bmatrix} \rho & \rho^2 & \dots & \rho^m \end{bmatrix}$, hvoraf det ses, at approksimationen for logdeterminanten for et bestemt ρ blot er givet ved det simple prikprodukt $\mathbf{a}^T\mathbf{b}$ mellem to vektorer af længde m, og dermed vil det være nemt at opdatere approksimationen for logdeterminanten for en ny værdi af ρ.

Man kan anvende egenskaben $\operatorname{tr}(W) = \sum_{i=1}^{m} \lambda_i \Rightarrow \operatorname{tr}(W^j) = \sum_{i=1}^{m} \lambda_i^j$ til at udregne sporet for potenser af W i (4.5), hvor det eneste der kræves er at finde egenværdierne λ_i for W.

4.2 ρ-estimatets følsomhed overfor approksimationsfejl

Når der arbejdes med estimering af ρ ved hjælp af approksimationer, er det ofte vigtig, at der er mulighed for at vurdere nøjagtigheden for de pågældende approksimationer. Denne nøjagtighed kan måles i forhold til påvirkningen på estimeringen af de pågældende variable, der er af interesse som for eksempel ρ. En metode kunne være, at se på hvordan uafhængige approksimationer af logdeterminanten påvirker den estimerede parameter ρ i SAR-modellen. Idet vi givet logdeterminantfunktionen hurtigt kan finde parameterestimatet for ρ, kan vi anvende m uafhængige approksimationer af logdeterminanten og således få m estimater for ρ. Variationen i disse estimater kan fungere som en indikation af nøjagtigheden for approksimationen af logdeterminanten. Hvis variationen er lille for parameterestimatet, indikerer dette, at approksimationsfejlen ikke forringer parameterestimationen betydeligt. Gode og i praksis anvendelige approksimationer har netop fejl, der ikke påvirker de pågældende resultater betydeligt.

Fejl i $\hat{\rho}$ ved brug af individuelle Monte Carlo-logdeterminant-approksimationer

I dette afsnit ser vi nærmere på spredningen for den estimerede værdi af ρ, der fremkommer ved approksimationen af logdeterminanten ved brug af Monte Carlo-approksimation. Det vises, at denne approksimation er yderst nøjagtig i den forstand, at den ikke betydeligt påvirker estimatet eller inferens med hensyn til parameteren ρ.

Eksempel 4.4:
Vi ser på et dataeksempel for at vurdere nøjagtigheden ved anvendelse af Monte Carlo-approksimation til at approksimere logdeterminanten. Vi laver et eksperiment med n observationer, hvor n varierer mellem 1.000 og 1.024.000. For hvert antal af observationer n genereres en stokastisk mængde af punkter, en vægtmatrix W, og den afhængige variabel $\mathbf{y} = (I_n - \rho W)^{-1}(X\boldsymbol{\beta} + \boldsymbol{\varepsilon})$ simuleres. Designmatricen X indeholder en skæring og en vektor bestående af normalfordelte stokastiske variable, og $\boldsymbol{\beta}$ er sat til ι_2. For $\boldsymbol{\varepsilon}$ er der anvendt en i.i.d. normalfordelt vektor med en standardafvigelse på 0,25. For hvert $\mathbf{y}$ er udregnet 100 individuelt approksimerede logdeterminanter ved brug af Monte Carlo-approksimation. For at vurdere hvor nøjagtigt en enkelt approksimation for logdeterminanten er, sammenlignes med et gennemsnit taget over 100 approksimationer for logdeterminanten, hvilket ses i figur 4.1.

n	$\tilde{\rho}_{\overline{\ln\lvert z\rvert_i}}$	mean $\tilde{\rho}_{\ln\lvert z\rvert_i}$	range $\tilde{\rho}_{\ln\lvert z\rvert_i}$
1,000	0.762549	0.762550	0.002119
2,000	0.752098	0.752099	0.001359
4,000	0.751474	0.751475	0.000873
8,000	0.752152	0.752152	0.000748
16,000	0.747034	0.747034	0.000485
32,000	0.748673	0.748673	0.000339
64,000	0.750238	0.750238	0.000224
128,000	0.750622	0.750622	0.000147
256,000	0.749766	0.749766	0.000122
512,000	0.750144	0.750144	0.000068
1,024,000	0.749404	0.749404	0.000052

Figur 4.1: Estimater for ρ baseret på gennemsnit og individuelle Monte Carlo-logdeterminant-approksimationer [LeSage and Pace, 2009, s. 101].

Det ses, at det estimerede $\hat{\rho}$ ikke er følsom overfor numerisk approksimationsfejl ved anvendelsen af individuelle approksimationer fra Monte Carlo-approksimation. Det ses ligeledes, at den estimerede parameter $\hat{\rho}$ varierer over et lille interval, som aftager med n. Estimationen af ρ over gennemsnittet er meget tæt på den rigtige værdi for ρ på 0,75. Når for eksempel n = 1.024.000, så er gennemsnitsestimatet for ρ over de 100 seperate log-determinanter lig med 0,749404, og forskellen på den største og den mindste estimerede værdi er 0,000052. Ved brug af gennemsnittet over de 100 log-determinant-estimater fås det samme estimat for ρ helt ned til seks decimaler. □

Nøjagtighed for approksimation af sporet

Vi har set, at nøjagtigheden ved Monte Carlo-approksimation er ret stor, og dette kan skyldes at metoden approksimerer sporet særdeles godt og/eller at loglikelihoodfunktionen ikke er følsom over for små variationer i logdeterminanten. Vi vil nu vurdere nøjagtigheden for approksimationen for sporet, så vi ser på den approksimerede vær-

di $\hat{\text{tr}} \approx n^{-1}\,\text{tr}(W^2)$ og sammenligner $\hat{\text{tr}}$ med den eksakte værdi for sporet $\text{tr} = n^{-1}\,\text{tr}(W^2)$.

Eksempel 4.5:
Der laves igen et eksperiment med n observationer, hvor n varierer fra 1.000 til 1.024.000, og der noteres 100 værdier ved hvert n. Det ses på figur 4.2, at Monte Carlo-approksimationen for sporet tilsyneladende virker nogenlunde middelværdiret, idet approksimationen har små gennemsnitlige fejl, der varierer i fortegn med n.

n	tr	average $\tilde{tr} - tr$	s.d. $\tilde{tr}$	range $\tilde{tr}$
1,000	0.166809	−0.000017	0.009841	0.046782
2,000	0.165991	0.000311	0.006742	0.034160
4,000	0.165782	−0.000296	0.004303	0.023378
8,000	0.165640	−0.000213	0.003138	0.018299
16,000	0.165505	0.000078	0.001957	0.009831
32,000	0.165420	0.000278	0.001690	0.009949
64,000	0.165391	0.000083	0.001050	0.005029
128,000	0.165395	0.000093	0.000799	0.003943
256,000	0.165377	0.000011	0.000634	0.003241
512,000	0.165359	−0.000023	0.000393	0.001820
1,024,000	0.165353	−0.000032	0.000277	0.001319

Figur 4.2: Individuelle spor estimeret ud fra MC af $n^{-1}\,\text{tr}\,W^2$ [LeSage and Pace, 2009, s. 102].

Desuden ses det, at standardafvigelsen og det interval, det estimerede spor befinder sig indenfor, er lille. For n = 16.000 ses en forskel på mindre end 0,01 mellem den største og den mindste værdi for det estimerede spor. Disse estimater er kun taget over 100 individuelle estimater, men hvis der laves et gennemsnit over alle de 100 forsøg ville fås et endnu mere nøjagtigt estimat. □

Teoretisk analyse af påvirkninger på $\hat{\rho}$ fra fejl i logdeterminatapproksimationer

Den autoregressive SAR-model $y = X\boldsymbol{\beta} + \rho W y + \boldsymbol{\varepsilon}$ har en profilloglikelihoodfunktion $L_p(\rho)$ som vist i (3.4), der involverer den rumligt afhængige parameter ρ, det vil sige, at

$$L_p(\rho) = \kappa + \ln\left|I_n - \rho W\right| - \frac{n}{2}\left(e(\rho)^T e(\rho)\right). \tag{4.7}$$

Hvis vi substituerer taylorudviklingen af logdeterminanten

$$ln\left|I_n - \rho W\right| = -\sum_{j=1}^{\infty} \frac{\rho^j\,\text{tr}(W^j)}{j}$$

i udtrykket i (4.7), fås at

$$L_p(\rho) = \kappa - \sum_{j=1}^{\infty} \frac{\rho^j\,\text{tr}(W^j)}{j} - \frac{n}{2}\left(e(\rho)^T e(\rho)\right). \tag{4.8}$$

Således har vi et udtryk for loglikelihoodfunktionen, der indeholder sporet, hvis rolle således kan analyseres.

For at analysere påvirkningen på (4.8) fra approksimationsfejl introduceres nu en skalarparameter δ_j til at modellere den proportionelle fejl ved approksimationen af $\text{tr}(W^j)$. Vi indsætter et led , der både afhænger af ρ og fejlen δ_j, der både kan være positiv eller negativ. Den nye profilloglikelihoodfunktion $L_p(\rho|\delta_j)$ for et givet δ_j bliver da

$$\begin{aligned} L_p(\rho|\delta_j) &= \kappa - \delta_j \frac{\rho^j \,\text{tr}(W^j)}{j} - \sum_{j=1}^{\infty} \frac{\rho^j \,\text{tr}(W^j)}{j} - \frac{n}{2}\left(e(\rho)^T e(\rho)\right) \\ &= -\delta_j \frac{\rho^j \,\text{tr}(W^j)}{j} + L_p(\rho) \\ &= \delta_j G(\rho) + L_p(\rho), \end{aligned} \tag{4.9}$$

hvor $G(\rho) = -\frac{\rho^j \,\text{tr}(W^j)}{j}$.

Nu ses på scoringsfunktionen for profilloglikelihoodfunktionen i (4.9), altså den førstordensafledede

$$S_p(\rho|\delta_j) = \frac{dL_p(\rho|\delta_j)}{d\rho} = \delta_j \frac{dG(\rho)}{d\rho} + \frac{dL_p(\rho)}{d\rho}. \tag{4.10}$$

Udtrykket $S_p(\rho|\delta_j)$ sættes lig nul, idet dette er førsteordensbetingelsen for at finde et eventuelt maksimum, og dette leder til en implicit funktion F, hvor

$$F = S_p(\rho|\delta_j) = 0.$$

Hvis profilloglikelihoodfunktionen er unimodal, det vil sige, at den kun har ét maksimum, er førsteordensbetingelsen ikke bare en nødvendig, men en tilstrækkelig betingelse for at sikre en unik løsning af ligningen $F = 0$ med hensyn til ρ. Hvis ikke der haves unimodalitet, kan der ses på den andenordensafledede for at sikre, at der er tale om et maksimum og ikke et minimum mellem to maksima.

Løsningen af ligningen afhænger af fejlen δ_j. Vi kan måle følsomheden af parameteren ρ med hensyn til fejlen δ_i ved at anvende implicitfunktionssætningen for to variable, så

$$\frac{d\rho}{d\delta_j} = -\frac{dF}{d\delta_j}\left(\frac{dF}{d\rho}\right)^{-1},$$

hvor det fra (4.10) ses, at

$$\frac{dF}{d\delta_j} = \frac{dG(\rho)}{d\rho} \qquad \text{og} \qquad \frac{dF}{d\rho} = \delta_j \frac{d^2G(\rho)}{d\rho^2} + \frac{d^2L_p(\rho)}{d\rho^2}.$$

Så får vi, at

$$\frac{d\rho}{d\delta_j} = -\frac{dG(\rho)}{d\rho}\left(\delta_j \frac{d^2G(\rho)}{d\rho^2} + \frac{d^2L_p(\rho)}{d\rho^2}\right)^{-1}. \tag{4.11}$$

Vi kan forsimple udtrykket i (4.11) ved $\frac{d^2 L_p(\rho)}{d\rho^2} = H(\rho)$, der er hessematricen, så

$$\frac{d\rho}{d\delta_j} = -\frac{dG(\rho)}{d\rho}\left(\delta_i \frac{d^2 G(\rho)}{d\rho^2} + H(\rho)\right)^{-1}.$$

Givet den estimerede værdi $\hat{\rho}$ er den estimerede varians for $\hat{\rho}$ givet ved $\hat{\sigma}^2(\hat{\rho})$. Vi kan nu anvende relationen mellem hessematricen og variansen for et parameterestimat for en univariat funktion såsom vores profilloglikelihood, så

$$\begin{aligned} \hat{\sigma}^2(\hat{\rho}) &= -H(\hat{\rho})^{-1} \Rightarrow \\ -\hat{\sigma}^{-2}(\hat{\rho}) &= H(\hat{\rho}). \end{aligned}$$

Vi har således, at

$$\frac{d\rho}{d\delta_j} = -\frac{dG(\rho)}{d\rho}\left(\delta_i \frac{d^2 G(\rho)}{d\rho^2} - \hat{\sigma}^{-2}(\hat{\rho})\right)^{-1}. \tag{4.12}$$

Hvis vi evaluerer (4.12) omkring punktet $\delta_i = 0$ fås, idet $G = \frac{\rho^j \operatorname{tr}(W^j)}{j}$, at

$$\begin{aligned} \left.\frac{d\rho}{d\delta_j}\right|_{\delta_i=0} &= -\frac{dG(\rho)}{d\rho}\left(0 - \hat{\sigma}^{-2}(\hat{\rho})\right)^{-1} \\ &= \frac{dG(\rho)}{d\rho}\hat{\sigma}^2(\hat{\rho}) \\ &= -(\hat{\rho})\rho^{j-1}\operatorname{tr}(W^j)\hat{\sigma}^2. \end{aligned}$$

Det ses dermed, at $\frac{d\hat{\rho}}{d\delta_j} < 0$ for positivt ρ, hvilket rent intuitivt giver mening, idet en positiv fejl, det vil sige $\delta_i > 0$, resulterer i en større logdeterminant-straf. Ved fejl vil værdien for $\hat{\rho}$ mindskes, således at $\hat{\rho}$ har en tendens til at flytte sig i den negative retning væk fra middelværdien. Desuden aftager den estimerede varians $\hat{\sigma}^2(\hat{\rho})$ ved stigende n, mens $\operatorname{tr}(W^j)$ vokser med stigende n, idet matricen således bliver større, og der dermed er flere tal, der lægges sammen, når man tager sporet af matricen. Dog skal det bemærkes, at dette ikke er en helt præcis antagelse, idet tallene i matricen muligvis vil ændre sig, når n vokser, og desuden kan det ikke sammenlignes med sporet for W, som jo er givet ved $\operatorname{tr}(W) = 0$, idet W har nul på diagonalen. Men mere generelt indikerer det, at produktet af de to variable ikke burde variere betydeligt.

Alt i alt afhænger $\hat{\rho}$'s følsomhed overfor approksimationsfejl i logdeterminanten i et enkelt spor af: Variansen for den estimerede $\hat{\rho}$, den sande værdi af sporet, den sande værdi af ρ og den proportionelle fejl ved approksimation af sporet. Kombinationen af de første tre faktorer burde resultere i en værdi, der ikke varierer betydeligt med n. Det skyldes, at sporet vokser lineært med n, mens variansen aftager lineært med n. Dog aftager den proportionelle fejl i approksimationen af sporet med n, og dermed har vi, at ρ's følsomhed overfor fejl i approksimationen af logdeterminanten aftager med antallet af observationer.

4.3 Udtryk for påvirkninger

Som før nævnt kræves relativt få udregninger til at finde udtrykket for sporet. Derfor ser vi nu på, hvordan sporet kan udnyttes i udregninger for gennemsnitlige påvirkninger.

Vi så i afsnit 2.5, at påvirkningen på den forventede værdi af den afhængige variabel, der skyldes ændringer i den r'te ikke-konstante baggrundsvariabel, er en funktion af matricen $S_r(W)$, så

$$E[\mathbf{y}] = \sum_{r=1}^{k} S_r(W)\mathbf{x}_r + \alpha^* \iota_n,$$

hvor $\alpha^* \iota_n$ repræsenterer skæringen, og k er antallet af søjler $\mathbf{x}_r$ fra designmatricen X, der summeres over.

Matricen $S_r(W)$ for SAR-modellen er som tidligere vist givet ved

$$S_r(W) = \left(I_n - \rho W\right)^{-1} \beta_r = I_n\beta_r + \rho W\beta_r + \rho^2 W^2\beta_r + \ldots,$$

og det samlede antal af uafhængige variable er $k+1$ for SAR-modellen, hvor k er antallet af parametre i $\boldsymbol{\beta}$, og 1-tallet kommer fra vektoren $\alpha\iota_n$. Vi så ligeledes i afsnit 2.5, at de gennemsnitlige påvirkninger ved inddragelse af matricen $S_r(W)$ er givet ved

$$\begin{aligned}
\bar{M}(r)_{direkte} &= n^{-1}\operatorname{tr}(S_r(W)) \\
\bar{M}(r)_{total} &= n^{-1}\iota_n^T S_r(W)\iota_n \\
\bar{M}_{indirekte} &= \bar{M}(r)_{total} - \bar{M}(r)_{direkte} = n^{-1}\iota_n^T S_r(W)\iota_n - n^{-1}\operatorname{tr}(S_r(W)).
\end{aligned} \tag{4.13}$$

Som tidligere forklaret er det ikke beregningsmæssigt effektivt at udregne påvirkningen direkte ud fra ovenstående ligninger via definitionen for $S_r(W)$, idet det ville involvere tæt besatte $(n \times n)$-matricer. Udfordringen ligger i beregningen af udtrykket $\operatorname{tr}(S_r(W))$, men som vi har set i afsnit 4.1, kan sporet approksimeres.

Sætning 4.6:
For SAR-modellen er de gennemsnitlige påvirkninger givet ved

$$\bar{M}_{direkte} = \boldsymbol{\beta} TG\iota_{q+1} \tag{4.14}$$

$$\bar{M}_{total} = \boldsymbol{\beta} g\iota_{q+1} \tag{4.15}$$

$$\bar{M}_{indirekte} = \boldsymbol{\beta} g\iota_{q+1} - \boldsymbol{\beta} TG\iota_{q+1},$$

hvor $T = \begin{bmatrix} 1 & 0 & n^{-1}\operatorname{tr}(W^2) & \ldots & n^{-1}\operatorname{tr}(W^q) \end{bmatrix}$, og $g = \begin{bmatrix} 1 & \rho & \rho^2 & \ldots & \rho^q \end{bmatrix}^T$, så $G_{ii} = g_i$ for $i = 1, \ldots, q+1$.

Bevis:
Lad T være en $(1 \times (q+1))$-vektor indeholdende de gennemsnitlige, diagonale elementer af potenser af W, så

$$\begin{aligned}
T &= \begin{bmatrix} n^{-1}\operatorname{tr}(W^0) & n^{-1}\operatorname{tr}(W^1) & n^{-1}\operatorname{tr}(W^2) & \ldots & n^{-1}\operatorname{tr}(W^q) \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & n^{-1}\operatorname{tr}(W^2) & \ldots & n^{-1}\operatorname{tr}(W^q) \end{bmatrix}.
\end{aligned}$$

Lad nu G repræsentere en $\big((q+1)\times(q+1)\big)$-diagonalmatrix, så $G_{ii} = g_i$ for $i = 1,\ldots,q+1$, hvor $g = \begin{bmatrix}1 & \rho & \rho^2 & \ldots & \rho^q\end{bmatrix}^T$, således at G indeholder potenser af ρ. Desuden haves vektoren $\boldsymbol{\beta} = \begin{bmatrix}\beta_1 & \beta_2 & \ldots & \beta_k\end{bmatrix}^T$.

Vi kan nu konstruere en række $(k\times 1)$-vektorer indeholdende den akkumulerede påvirkning for hver af de $r = 1,\ldots,k$ ikke-konstante baggrundsvariable, så (4.14) og (4.15) fremkommer. Nu bevises, at de gennemsnitlige påvirkninger givet ved disse udtryk er de samme som i (4.13).

Ligheden i (4.14) kan udregnes til

$$\begin{aligned}
\bar{M}_{direkte} &= \boldsymbol{\beta} T G \iota_{q+1} \\
&= \begin{bmatrix}\beta_1\\ \vdots \\ \beta_k\end{bmatrix}\begin{bmatrix}n^{-1}\operatorname{tr}(W^0) & \ldots & n^{-1}\operatorname{tr}(W^q)\end{bmatrix}\begin{bmatrix}1 & 0 & \cdots & 0\\ 0 & \rho & \ddots & \vdots\\ \vdots & \ddots & \ddots & 0\\ 0 & \cdots & 0 & \rho^q\end{bmatrix}\begin{bmatrix}1\\1\\\vdots\\1\end{bmatrix} \\
&= \begin{bmatrix}n^{-1}\operatorname{tr}(W^0)\beta_1 & \ldots & n^{-1}\operatorname{tr}(W^q)\beta_1\\ \vdots & \vdots & \vdots \\ n^{-1}\operatorname{tr}(W^0)\beta_k & \ldots & n^{-1}\operatorname{tr}(W^q)\beta_k\end{bmatrix}\begin{bmatrix}1\\ \rho\\ \vdots\\ \rho^q\end{bmatrix} \\
&= \begin{bmatrix}n^{-1}\operatorname{tr}(W^0)\beta_1 + \rho n^{-1}\operatorname{tr}(W)\beta_1 + \ldots + \rho^q n^{-1}\operatorname{tr}(W^q)\beta_1\\ \vdots \\ n^{-1}\operatorname{tr}(W^0)\beta_k + \rho n^{-1}\operatorname{tr}(W)\beta_k + \ldots + \rho^q n^{-1}\operatorname{tr}(W^q)\beta_k\end{bmatrix} \\
&= n^{-1}\sum_{j=0}^{q}\rho^j \operatorname{tr}(W^j)\boldsymbol{\beta}.
\end{aligned}$$

Her ses det ved sammenligning med (2.16), at vi får det samme som det oprindelige udtryk for den gennemsnitlige, direkte påvirkning.

Ligheden i (4.15) kan udregnes til

$$\begin{aligned}
\bar{M}_{total} &= \boldsymbol{\beta} g \iota_{q+1} \\
&= \boldsymbol{\beta} \sum_{j=0}^{q} \rho^j \\
&= \sum_{j=0}^{q} \rho^j \boldsymbol{\beta} \\
&= n^{-1} \sum_{j=0}^{q} \rho^j n \boldsymbol{\beta} \\
&= n^{-1} \sum_{j=0}^{q} \rho^j \iota_n^T W^j \iota_n \boldsymbol{\beta} \\
&= n^{-1} \iota_n^T \sum_{j=0}^{q} \rho^j W^j \iota_n \boldsymbol{\beta} \\
&= n^{-1} \iota_n^T \left(I_n - \rho W\right)^{-1} \iota_n \boldsymbol{\beta},
\end{aligned}$$

idet W er en rækkestokastisk $(n \times n)$-matrix. Her ses det ved sammenligning med (2.17), at vi får det samme som det oprindelige udtryk for den gennemsnitlige, samlede påvirkning.

Den indirekte gennemsnitlige påvirkning fås da ved

$$\bar{M}_{indirekte} = \bar{M}_{total} - \bar{M}_{direkte} = \boldsymbol{\beta} g \iota_{q+1} - \boldsymbol{\beta} T G \iota_{q+1}.$$

■

Man kan ligeledes se på estimering af grupperede rumlige påvirkninger, der siger noget om, hvordan påvirkningerne aftager, når vi bevæger os længere og længere væk mod naboer af højere orden. Vektoren ι_{q+1} akkumulerer påvirkningerne over alle potenserne 0 til q ved brug af matricen $(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \ldots$. Hvis vi erstatter denne $((q+1) \times 1)$-vektor med en vektor indeholdende 1 i den r'te indgang og 0 i resten af indgangene, fås påvirkningerne for SAR-modellen til

$$\begin{aligned}
\bar{M}(r)_{direkte} &= \beta_r \\
\bar{M}(r)_{total} &= \beta_r \\
\bar{M}(r)_{indirekte} &= \beta_r - \beta_r = 0.
\end{aligned}$$

Således kan man se på effekter fra naboer af forskellige ordener ud fra $\bar{M}$ ved at skifte ι_n ud med en vektor med 1 placeret i den r'te indgang og nul på resten af indgangene, eller hvis man ønsker decideret at udelade påvirkninger fra specifikke naboer. Hvis vi for eksempel sætter de to første indgange i vektoren ι_{q+1} til 1 og resten sættes til 0, fås estimater for påvirkningen fra nulteordensnaboer adderet med påvirkningen fra førsteordensnaboer. Hvis man ønsker at se på alle påvirkninger op til en specifik orden t, sættes de første t indgange i $((q+1) \times 1)$-vektoren ι_{q+1} lig med 1 og de restende lig med 0. Således fås en rumlig opdeling af de direkte påvirkninger, de indirekte påvirkninger og den samlede påvirkning.

4.4 Lukkede løsninger for rumlige modeller med én parameter

Førsteordensbetingelsen, at den førsteordensafledede skal være lig med nul for at maksimere profilloglikelihoodfunktionen, leder til et polynomium, der indeholder ρ som eneste parameter. Problemet har en løsning på lukket form i den forstand, at man kan udtrykke den optimale værdi for den pågældende parameter ved hjælp af egenværdier for en lille matrix.

Sætning 4.7:
En polynomisk approksimation af logdeterminanten for SAR-modellen er givet ved

$$\ln\left|I_n - \rho W\right| \approx \mathbf{c}^T \mathbf{v}(\rho),$$

hvor $\mathbf{v}(\rho) = \begin{bmatrix}\rho & \rho^2 & \dots & \rho^q\end{bmatrix}^T$, og $\mathbf{c} = \begin{bmatrix}-\operatorname{tr}(W) & -\frac{1}{2}\operatorname{tr}(W^2) & \dots & -\frac{1}{q}\operatorname{tr}(W^q)\end{bmatrix}$.

Bevis:
Hvis vi udregner $\mathbf{c}^T\mathbf{v}(\rho)$ fås, at

$$\mathbf{c}^T\mathbf{v}(\rho) = \begin{bmatrix}-\operatorname{tr}(W) & -\frac{1}{2}\operatorname{tr}(W^2) & \dots & -\frac{1}{q}\operatorname{tr}(W^q)\end{bmatrix}\begin{bmatrix}\rho\\ \rho^2\\ \vdots\\ \rho^q\end{bmatrix} = -\sum_{j=1}^{k}\frac{\rho^j \operatorname{tr}(W^j)}{j},$$

hvilket er identisk med udtrykket i (4.6), så det haves, at

$$\mathbf{c}^T\mathbf{v}(\rho) \approx \ln\left|I_n - \rho W\right|.$$

■

For SAR-modellen kan vi opsætte det kvadratiske polynomium for det kvadrerede fejlled, så

$$S(\rho) = e(\rho)^T e(\rho) = \mathbf{u}(\rho)^T Q \mathbf{u}(\rho),$$

hvor

$$\begin{aligned} e(\rho) &= \left(I_n - X(X^TX)^{-1}X^T\right) Y\mathbf{u}(\rho)\\ \mathbf{u}(\rho) &= \begin{bmatrix}1 & -\rho\end{bmatrix}^T\\ Q &= Y^T\left(I_n - X(X^TX)^{-1}X^T\right)Y\\ Y &= \begin{bmatrix}\mathbf{y} & W\mathbf{y}\end{bmatrix}. \end{aligned}$$

Ved brug af (3.4) fås nu profilloglikelihoodfunktionen

$$\begin{aligned} L_p(\rho) &= \kappa + \ln\left|I_n - \rho W\right| - \frac{n}{2}\left(S(\rho)\right)\\ &\approx \kappa + \mathbf{c}^T\mathbf{v}(\rho) - \frac{n}{2}\left(\mathbf{u}(\rho)^T Q\mathbf{u}(\rho)\right). \end{aligned}$$

Nu kan den førsteordensafledede findes som

$$\frac{dL_p(\rho)}{d\rho} = \frac{d\ln\left|I_n - \rho W\right|}{d\rho} - \frac{n}{2}S(\rho)^{-1}\frac{dS(\rho)}{d\rho},$$

og hvis denne sættes lig nul og alle led multipliceres med $S(\rho)$, fås det at

$$\frac{d\ln\left|I_n - \rho W\right|}{d\rho}S(\rho) - \frac{n}{2}\frac{dS(\rho)}{d\rho} = 0.$$

Hvis vi multiplicerer ovenstående med n^{-1}, fås en ækvivalent førsteordenbetingelse, som har bedre numeriske egenskaber for stort n, så

$$\frac{1}{n}\frac{d\ln\left|I_n - \rho W\right|}{d\rho}S(\rho) - \frac{1}{2}\frac{dS(\rho)}{d\rho} = 0. \tag{4.16}$$

Det første led i (4.16) er et produkt af to størrelser. Den ene er det positivt definitte kvadratiske udtryk $S(\rho)$, som er et andengradsolynomium med hensyn til ρ, hvis koefficienter er summen langs antidiagonalen af Q. Den anden størrelse er den afledede af logdeterminanten, som er et polynomium af grad $p-1$. Ved multiplicering af de to størrelser haves altså et polynomium af grad $p+1$. Det andet led består af den afledede af $S(\rho)$, som er et førstegradpolynomium. Hvis udtrykket i (4.16) udregnes, haves et samlet polynomium

$$\begin{aligned} &P(\rho) = a_0 + a_1\rho + \ldots + a_p\rho^p + a_{p+1}\rho^{p+1} \Rightarrow \\ &\frac{P(\rho)}{a_{p+1}} = c_0 + c_1\rho + \ldots + c_p\rho^p + \rho^{p+1} \qquad \text{for} \qquad c_i = \frac{a_i}{a_{p+1}} \end{aligned}$$

af grad $p+1$, som dermed har $p+1$ mulige løsninger. Polynomiet kan opskrives på matrixform, så

$$C(P) = \begin{bmatrix} 0 & 0 & \ldots & 0 & -c_0 \\ 1 & 0 & \ldots & 0 & -c_1 \\ 0 & 1 & \ddots & \vdots & -c_2 \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \ldots & 0 & 1 & -c_p \end{bmatrix},$$

der er en $\big((p+1)\times(p+1)\big)$-matrix, hvor de $p+1$ løsninger til ligningssytemet i (4.16) er givet ved egenværdierne for matricen $C(P)$. Blandt de $p+1$ løsninger givet ved egenværdierne findes den værdi for ρ, der maksimerer likelihoodfunktionen.

Her findes der således en løsning på lukket form, der udtrykkes ved egenværdier for en lille matrix. Beregningen af egenværdier i dette tilfælde kræver $O((p+1)^3)$ operationer og afhænger således ikke af n, hvilket gør metoden anvendelig i praksis for store, rumlige datasæt.

Modeltest 5

Dette kapitel er skrevet på baggrund af [LeSage and Pace, 2009, kapitel 3], [Azzalini, 1996, afsnit 3.3.3 og kapitel 4], [Olofsson, 2005, afsnit 6.3 og 6.5-6.6].

Modeltest og inferens er en vigtig del af arbejdet med at tilpasse modeller til data. Man er oftest interesseret i at kunne fortolke modellen og dens parametre efterfølgende, således at der kan drages konklusioner om data på baggrund af den anvendte model. Ved anvendelse af maximum likelihood-baserede inferensmetoder sikres det ved determinantleddet, at estimatet $\hat{\rho}$ er i intervallet, der er defineret af vægtmatricens maksimale og minimale egenværdi, hvilket for SAR-modellen er $(\lambda_{\min}^{-1},1)$. Derfor fokuseres nu på likelihoodbaserede inferensmetoder, hvor der indledes med et kort afsnit om hypotesetest og tilhørende begreber, hvorefter der ses på forskellige test.

5.1 Hypotesetest

Når man har opstillet en model for en række data, er man ofte interesseret i at teste parametrene i modellen og dermed undersøge en eller anden form for hypotese for de data, som modellen beskriver.

Definition 5.1 (Teststørrelse):
En teststørrelse er en funktion $T(\mathbf{y}) : \mathbb{R}^n \to \mathbb{R}$, hvor $T(\mathbf{y})$ er den observerede værdi.

Teststørrelsen har således en fast værdi og kaldes den observerede værdi. Den kan anvendes til at teste følgende hypotese

$$\begin{cases} H_0: & \theta_i \in \Theta_0 \\ H_1: & \theta_i \in \Theta_1, \end{cases}$$

hvor $\Theta_0,\Theta_1 \subseteq \mathbb{R}$, og θ_i er et element i parametervektoren $\boldsymbol{\theta}$. H_0 kaldes for *nulhypotesen* og H_1 for *den alternative hypotese*. Man ønsker ofte at teste hypotesen

$$\begin{cases} H_0: & \theta_i \in \Theta_0 = \{K\} \\ H_1: & \theta_i \in \Theta_1 = \mathbb{R}\backslash\{K\}, \end{cases}$$

hvor konstanten K i mange tilfælde er 0.

Da θ_i er ukendt, vides det ikke med sikkerhed, om den ligger i H_0 eller ej, og der vil altid være risiko for at begå én ud af to typer af fejl. Man kan acceptere en falsk nulhypotese, eller vi kan forkaste en sand (se tabel 5.1). Når en nulhypotese testes, skal det således besluttes, hvor stor en fejlsandsynlighed der kan tolerere.

	H_0 accepteres ($\theta_i \in \Theta_0$)	H_0 forkastes ($\theta_i \notin \Theta_0$)
H_0 sand ($\theta_i \in \Theta_0$)	Korrekt	Fejl af type 1
H_0 falsk ($\theta_i \notin \Theta_0$)	Fejl af type 2	Korrekt

Tabel 5.1: Fejltyper ved test af nulhypoteser.

Hypotesetest afspejler det faktum, at det er nemmere at forkaste en påstand end at bevise den. Når en omstændighed testes, er det derfor oftest den alternative hypotese, det ønskes at kunne bekræfte eller i hvert fald ikke forkaste. For eksempel kan man for SAR-modellen ønske at vise, at der er rumlig afhængighed til stede i data. I dette tilfælde antages den nulhypotese, at der ikke er rumlig afhængighed, hvoraf den alternative hypotese er, at der er rumlig afhængighed i data. Det er klart, at det ønskes, at nulhypotesen forkastes til fordel for, at den alternative hypotese virker rimelig. Hvis nulhypotesen ikke kan forkastes, så accepteres den, men det betyder ikke, at den er bevist, det betyder blot, at vi ifølge data ikke kan forkaste den.

Signifikansniveau og *p*-værdi

I det tilfælde, hvor der kan bestemmes en fordeling for teststørrelsen $T(\mathbf{Y})$, kan der fastsættes et signifikansniveau og bestemmes en p-værdi, således at den stokastiske teststørrelse $T(\mathbf{Y})$ kan sammenlignes med den observerede værdi i teststørrelsen $T(\mathbf{y})$.

Definition 5.2 (Signifikansniveau):
Et signifikansniveau α er givet ved

$$\alpha = \sup_{\theta_i \in \Theta_0} P_{\theta_i}(T(\mathbf{Y}) \in \Theta_1),$$

hvor Θ_1 er det kritiske område.

Det kritiske område er ofte en fastsat værdi eller et interval som for eksempel $\Theta_1 = (-\infty, k]$, hvor k kaldes den kritiske værdi. Det kræves, at fordelingen under nulhypotesen er kendt. Signifikansniveauet er hermed lig med sandsynligheden for at forkaste en nulhypotese, selvom den er sand, hvilket er det samme som sandsynligheden for at begå en type 1-fejl. Oftest vælges standardværdier som 0,05, 0,01 eller 0,001 svarende til henholdsvis 5%, 1% eller 0,1% for α. Følgende fremgangsmåde anvendes ved hypotesetest ved hjælp af signifikansniveauer.

1. Fastsæt en nulhypotese H_0 og en alternativ hypotese H_1.
2. Find teststørrelsen $T(\mathbf{Y})$ og bestem for hvilke værdier (små, store, positive, negative) nulhypotesen skal forkastes til fordel for den alternative hypotese. Dette kan lade sig gøre under den forudsætning, at fordelingen for $T(\mathbf{Y})$ er helt kendt under antagelsen at nulhypotesen H_0 er sand.

3. Vælg et signifikansniveau $\alpha = \sup_{\theta_i \in \Theta_0} P_{\theta_i}(T(\mathbf{Y}) \in \Theta_1)$, og find det kritiske område Θ_1 ved at antage, at H_0 er sand. Her fås de faktiske tal beskrevet generelt i punkt 2.
4. Udregn $T(\mathbf{Y})$ og sammenlign med det kritiske område. Hvis $\mathbf{Y} \in \Theta_1$ forkastes nulhypotesen H_0 til fordel for den alternative hypotese H_1, og hvis ikke accepteres H_0.

Valget af signifikansniveau α kan være en smule arbitrært på trods af, at bestemte værdier som ovenfor nævnt anses for at være typiske, og desuden er det ikke altid tilfredsstillende at vælge signifikansniveauet på forhånd. For eksempel kan det være, at man kan forkaste et nulhypotese ved et endnu lavere signifikansniveau end det, der på forhånd er fastsat, således at afvisningen hermed bliver stærkere. Man kan have afvist en nulhypotese med fastsat signifikansniveau på 0,05, men som viser sig at kunne være afvist med et signifikansniveau på 0,01, således at der, i stedet for at være 5% sandsynlighed for at forkaste en sand nulhypotese, kun er 1% sandsynlighed for dette. Der kan således ses på det laveste signifikansniveau, for hvilket nulhypotesen H_0 for data kan forkastes. Ved at angive værdien for det laveste signifikansniveau, hvor nulhypotesen kan forkastes, haves et mål for, hvor stærkt et bevis data giver imod H_0.

Definition 5.3 (p-værdi):
En p-værdi for θ_i er givet ved

$$p = \sup_{\theta_i \in \Theta_0} P_{\theta_i}\left(|T(\mathbf{Y})| > |T(\mathbf{y})|\right),$$

hvor $T(\mathbf{y})$ er den observerede værdi og $\mathbf{Y}$ er er en stokastisk variabel.

Det vil sige, at en p-værdi for en test er det laveste signifikansniveau på hvilket nulhypotesen H_0 kan forkastes for et givet datasæt. Hvis der haves en p-værdi p for givet data, så afvises nulhypotesen på signifikansniveauet $p \leq \alpha$. Værdien angiver så at sige sandsynligheden for, at en teststørrelse $T(\mathbf{Y})$ giver en mere ekstrem værdi end den observerede værdi $T(\mathbf{y})$ under antagelse af, at nulhypotesen er sand. En værdi a siges at være mere ekstrem end en anden værdi b, hvis a er mere kritisk for nulhypotesen end b.

Hvis man eksempelvis har at gøre med SAR-modellen og ønsker at teste, om der er rumlig afhængighed, så antages det modsatte, det vil sige, at nulhyposen er, at den rumlige afhængighed er nul. Det er klart, at denne hypotese ønskes forkastet, men man er også oftest interesseret i, at der er en lille sandsynlighed for at begå en fejl af type 1. Der ønskes derfor lave værdier for p. Ligeledes vælges α til at være lav, men den skal heller ikke vælges for lav, da der i så fald risikeres, at der er mange fejl af type 2.

Eksempel 5.4:
Hvis $T(\mathbf{Y})$ har en kendt t-fordeling med tæthedsfunktion f, og både store og små værdier antages at være kritiske for nulhypotesen, haves det, at

$$p = \sup_{\theta_i \in \Theta_0} P_{\theta_i}\left(|T(\mathbf{Y})| > |T(\mathbf{y})|\right) = \sup_{\theta_i \in \Theta_0} \left(1 - \int_{-T(\mathbf{y})}^{T(\mathbf{y})} f(\mathbf{y})\, d\mathbf{y}\right),$$

da en t-fordeling er symmetrisk omkring nul. For en t-fordelt teststørrelse er acceptområdet givet ved $\Theta_0 = [-t_{\alpha/2}, t_{\alpha/2}]$. Værdien t_α bestemmes ved at løse

$$\alpha = \sup_{\theta_i \in \Theta_0} P_{\theta_i}(T(\mathbf{Y}) \in \Theta_1) = \sup_{\theta_i \in \Theta_0} (1 - P(T(\mathbf{Y}) \in \Theta_0)) = \sup_{\theta_i \in \Theta_0} \left(1 - \int_{-t_{\alpha/2}}^{t_{\alpha/2}} f(\mathbf{y}) d\mathbf{y}\right).$$

Da α er valgt på forhånd, kan ovenstående ligning løses. □

5.2 Konfindensinterval

Når man arbejder med estimationer af parametre i en model, er det oftest nyttigt at kunne sige noget om, hvor præcise parameterestimeringerne er. Til dette formål anvendes konfidensintervaller, der er en metode til at angive hvor stor en usikkerhed, der er forbundet med estimationen af modellens parametre.

Definition 5.5 (Konfidensinterval):
Et konfidensniveau er givet ved

$$\alpha = P(T_1(Y) < \theta_i < T_2(Y)),$$

hvor T_1 og T_2 er to funktioner af den stokastiske variabel Y, og θ_i er den parameter, der ønskes et konfidensinterval for. Intervallet $[T_1(Y), T_2(Y)]$ kaldes konfidensintervallet for θ_i med konfidensniveau α.

Der er hermed tale om et interval, der indeholder θ_i med sandsynligheden α. Når der haves observerede værdier fra en fordeling over data, så haves numeriske værdier for $T_1(Y)$ og $T_2(Y)$, og disse siges at angive et observeret konfidensinterval.

Konfidensintervaller er ofte på formen $\left[\hat{\theta}_i - R, \hat{\theta}_i + R\right]$, hvor $\hat{\theta}_i$ er den estimerede parameter for θ_i, og R angiver en grænse, således at

$$\alpha = P\left(\hat{\theta}_i - R < \theta_i < \hat{\theta}_i + R\right).$$

I praksisk anvendes ofte metoden, hvor der først defineres en pivotstørrelse, der både afhænger af data $\mathbf{y}$ og den modelparameter θ_i, som ønskes et konfidensinterval for. Pivotstørrelsen skal have en kendt fordeling. Derefter fastsættes α, så

$$\alpha = P\left(q_L < \text{pivotstørrelse} < q_U\right),$$

hvor q_L er $\frac{1-\alpha}{2}$-fraktilen, og q_U er $1 - \frac{1-\alpha}{2}$-fraktilen. Hvis der ønskes konfidensintervaller for σ^2 og β_i, kunne pivotstørrelsen for eksempel være $\hat{\sigma}^2/\sigma^2$ eller $\hat{\beta}_i - \beta_i$, hvis disse elementers fordelinger er kendt.

Ved simpel omregning fås herefter, at

$$\alpha = P\left(q_L < \text{pivotstørrelse} < q_U\right) = P(T(Y) < \theta_i < T(Y)), \tag{5.1}$$

hvor parameteren θ_i isoleres, som det ses i (5.1), i midten af den første ulighed, og dermed fås konfidensintervallet for θ_i. Det gælder hermed, at procentdelen α af sandsynlighedsmassen er inden for det interval som q_L- og q_U-fraktilerne angiver. Oftest bestemmes α-konfidensintervaller med $\alpha = 95\%$, hvilket vil sige, at $q_L = 0{,}025$ og $q_U = 0{,}975$.

Konfidensintervallet afhænger af fordelingen for parameteren. Hvis der haves en normalfordeling, fås det approksimative konfidensinterval i definition 5.6.

Definition 5.6:
Hvis der simuleres n værdier for θ_i fra normalfordelingen $N(\mu,\sigma^2)$, og $\hat{\theta}_i$ er maximum likelihood-estimatet for disse, så er konfidensintervallet med konfidensniveau $\alpha = 0{,}95$ for μ givet ved

$$\mu = \hat{\theta}_i \pm z\frac{\sigma}{\sqrt{n}},$$

hvor z er 97,5-fraktilen for normalfordelingen, μ er en ukendt middelværdi og σ er standardafvigelsen.

Hvis der anvendes den estimerede værdi $\hat{\sigma}$, fås en t-fordeling i stedet for en normalfordeling, og de kritiske værdier for t-fordelingen indsættes på z's plads i (5.6). Ved den centrale grænseværdisætning går fordelinger inklusive t-fordelingen mod normalfordelingen, når antallet af målinger n går mod uendelig.

Det er også muligt at finde et konfidensinterval for σ^2, når μ er ukendt. Hvis der haves n simulerede værdier for θ_i fra $N(\mu,\sigma^2)$, hvor $\hat{\theta}_i$ som før er maximum likelihood-estimatet for disse, så gælder det ifølge [Olofsson, 2005, s. 317], at

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{n-1}, \tag{5.2}$$

hvor $n-1$ er antallet af frihedsgrader (se appendiks B.2). Resultatet i (5.2) kan anvendes til at finde et konfidensniveau α ved først at finde to værdier x_1 og x_2, der opfylder

$$\begin{aligned} &P\left(x_1 \leq \frac{(n-1)\hat{\sigma}^2}{\sigma^2} \leq x_2\right) = \alpha \qquad \Rightarrow \\ &P\left(\frac{(n-1)\hat{\sigma}^2}{x_2} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{x_1}\right) = \alpha. \end{aligned} \tag{5.3}$$

Der findes uendeligt mange måder at vælge x_1 og x_2 på, så de opfylder kravet i (5.3). Derfor vælges nu det ekstra krav, at intervallet skal være symmetrisk, så

$$P\left(x_1 \geq \frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right) = P\left(x_2 \leq \frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right) = \frac{1-\alpha}{2}.$$

Idet $1 - \frac{1-\alpha}{2} = \frac{1+\alpha}{2}$, fås konfidensintervallet givet i definition 5.7.

Definition 5.7:
Hvis der simuleres n værdier for θ_i fra normalfordelingen $N(\mu,\sigma^2)$, og $\hat{\theta}_i$ er maximum likelihood-estimatet for disse, så er et symmetrisk konfidensinterval med konfidensniveau $\alpha = 0{,}95$ for σ^2 givet ved

$$\frac{(n-1)\hat{\sigma}^2}{x_2} \leq \sigma^2 \leq \frac{(n-1)\hat{\sigma}^2}{x_1},$$

hvor μ er en ukendt middelværdi og $\hat{\sigma}$ er estimatet for standardafvigelsen. Desuden er $F_{\chi^2_{n-1}}(x_1) = \frac{1-\alpha}{2}$ og $F_{\chi^2_{n-1}}(x_2) = \frac{1+\alpha}{2}$, hvor $F_{\chi^2_{n-1}}(x_1)$ er fordelingsfunktionen for χ^2-fordelingen med $n-1$ frihedsgrader evalueret i punktet x_1.

5.3 Likelihood ratio test

Hvis man vil teste en nulhypotese kan man undersøge, hvor stor likelihoodfunktionen er i et punkt $\hat{\theta}_0$, der maksimerer likelihoodfunktionen under H_0, hvor indekset indikerer, at vi maksimerer $l(\theta_i|y)$ over Θ_0, hvor y er givet data. Der ses nu på forholdet mellem $l(\hat{\theta}_0|y)$ og maksimum likelihood-estimatet over hele Θ, så

$$LR = \frac{l(\hat{\theta}_0|y)}{l(\hat{\theta}_i|y)},$$

hvor $\hat{\theta}_i$ er den estimerede værdi for θ_i, der maksimerer likelihoodfunktionen.

Det bemærkes, at $0 < LR \leq 1$, og det ses desuden, at LR opfylder definition 5.1, og dermed er LR en teststørrelse. Hvis LR er tæt på 1, betyder det, at vores parameterestimat er tæt på at gøre likelihoodfunktionen maksimal, og vi er tilbøjelige til at acceptere nulhypotesen. Hvis LR er tæt på 0, er der grund til at tvivle på nulhypotesen.

I praksis udnytter man ofte, at der asymptotisk gælder, at

$$\begin{aligned} Q &= -2\left(\log\left(l(\hat{\theta}_0|y)\right) - \log\left(l(\hat{\theta}_i|y)\right)\right) \\ &= -2\left(L(\hat{\theta}_0|y) - L(\hat{\theta}_i|y)\right) \sim \chi^2_{(k+1)-d}, \end{aligned}$$

hvor $k+1$ er antallet af parametre i den fulde model, og d er antallet af parametre, der stadig varierer i modellen under nulhypotesen. Hermed er $k+1-d$ det antal parametre, der er sat lig konstanter eller nul i nulhypotesen, således at disse ikke længere ses som parametre i modellen. Bemærk, at her er store værdier for Q kritiske.

Ved maksimum likelihood-inferens anvendes ofte likelihood ratio-test til undersøgelse af for eksempel nulhypotesen for parametre i en anvendt model. Beregningstiden for likelihood ratio-test er hurtig, idet det at udregne likelihoodfunktioner ikke er særlig tidskrævende, så derfor kan denne test med fordel anvendes til for eksempel hypoteser om udeladelse af en enkelt baggrundsvariabel. I forbindelse med rumlige, økonometriske modeller kunne det være interessant at se på, hvorvidt der overhovedet er tale om rumlig afhængighed i de afhængige variable eller om for eksempel den rumligt afhængige parameter θ_i i SAR-modellen lige så godt kunne være lig med nul.

5.4 Waldtest

Waldtesten anvender hessematricen H enten numerisk eller analytisk for at få en varians-kovarians-matrix for de estimerede parametre og hermed t-testen. Vi vil fokusere på anvendelsen af hessematricen, hvis indgange består af andenordensafledede af loglikelihoodfunktionen med hensyn til alle relevante parametre i modellen. Men først introduceres selve Waldtesten, der bygger på antagelsen om, at maximum likelihood-estimater er asymptotisk normalfordelte med en kovariansmatrix, der er lig med den inverse Fisherinformation (se [Azzalini, 1996, s. 82-83]). Fisherinformationen er defineret ud fra scoringsfunktionen og er udtrykt i definition 5.8.

Definition 5.8 (Fisherinformation):
Fisherinformationen er defineret ved matricen

$$I(\boldsymbol{\theta}) := -\mathbf{E}\left[\frac{d}{d\boldsymbol{\theta}^T} s(\boldsymbol{\theta}|\mathbf{Y})\right] = -\mathbf{E}\left[\frac{d^2}{d\boldsymbol{\theta}\, d\boldsymbol{\theta}^T} L(\boldsymbol{\theta}|\mathbf{Y})\right],$$

hvor $s(\boldsymbol{\theta}|\mathbf{Y}) := \frac{d}{d\boldsymbol{\theta}} L(\boldsymbol{\theta}|\mathbf{Y})$, og middelværdien tages med hensyn til observationen Y_i, og $\boldsymbol{\theta}$ er en parameter, der holdes fast.

Fisherinformationen tager middelværdien for alle mulige værdier for Y_i.

Ved Waldtesten testes hypotesen

$$\begin{cases} H_0: & \theta_i = \theta_0 \\ H_1: & \theta_i \neq \theta_0. \end{cases}$$

Da den sande værdi for θ_i ikke kan udregnes, anvendes maximum likelihood-estimatet for θ_i givet ved $\hat{\theta}_i$ for at være i stand til at udføre testen, og det antages, at $\hat{\theta}_i - \theta_0$ er normalfordelt. Her arbejdes med teststørrelsen

$$T(\mathbf{y}) = \frac{\left(\hat{\theta}_i - \theta_0\right)^2}{\mathrm{Var}\left(\hat{\theta}\right)},$$

hvor vi har trukket den antagede middelværdi θ_0 fra maximum likelihood-estimatet og divideret med variansen for maximum likelihood-estimatet. Det testes om denne teststørrelse følger en χ^2-fordeling, det vil sige

$$T(y) = \frac{\left(\hat{\theta}_i - \theta_0\right)^2}{\mathrm{Var}\left(\hat{\theta}\right)} \sim \chi^2_\nu,$$

hvor $\mathrm{Var}\left(\hat{\theta}\right) \approx I^{-1}(\theta_i)$, og ν er lig antallet af de parametre, der er fastholdt af nulhypotesen. En sådan test kan overføres til tilfældet, hvor flere parametre ønskes testet, så

$$T(y) = \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)^T \mathrm{Var}(\hat{\boldsymbol{\theta}})^{-1} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \sim \chi^2_\nu,$$

hvor $\text{Var}\left(\hat{\boldsymbol{\theta}}\right) \approx I^{-1}\left(\boldsymbol{\theta}\right)$.

Man kan også opstille en lignende teststørrelse for én variabel, hvor det undersøges om

$$T\left(y\right) = \frac{\hat{\theta}_i - \theta_0}{\text{se}\left(\hat{\theta}\right)} \sim N\left(0,1\right),$$

og hvis ovenstående rent faktisk følger en standard normalfordeling, kan nulhypotesen $\theta_i = \theta_0$ accepteres. I forbindelse med denne teststørrelse anvendes, at $\text{se}\left(\hat{\theta}\right) \approx \sqrt{I^{-1}\left(\hat{\theta}_i\right)}$, og det ses i det følgende afsnit, at vi kan opfatte $I^{-1} = -H^{-1}$ som et estimat for kovariansmatricen, idet $I = -\,\text{E}\left[H\right]$, hvor H er hessematricen.

Nu ses nærmere på hessematricen H.

Kombineret analytisk-numerisk beregning af Hessematricen

Det kan være kompliceret at konstruere og udregne hessematricen, og der er styrker og svagheder både ved brugen af den analytiske hessematrix $H^{(a)}$ og den numeriske hessematrix $H^{(n)}$, som kan defineres på flere måder, til approksimation af hessematricen H. Når likelihoodfunktionen er kompliceret at differentiere, approksimeres den numeriske hessematrix. Det gøres ved at lave tangenter mellem punkterne på likelihoodfunktionen og anvende hældningerne til en kurve, for hvilken tangenterne udgør indgangene i hessematricen. De fleste elementer i den analytiske hessematrix tager ikke lang tid at udregne og er ikke så følsomme over for skaleringsproblemer. Men hvis den analytiske hessematrix skal udregnes direkte, kræves en beregning af sporet, der indeholder den inverse af den tæt besatte $(n \times n)$-matrix $(I_n - \rho W)^{-1}$, hvilket kan være kompliceret. Ved den numeriske tilgang drages der nytte af, at det er hurtigt at beregne loglikelihoodfunktionen, og den tager derfor mindre tid og er yderst anvendelig for det enkelte komplicerede element i den analytiske hessematrix. Det er derfor interessant at se på, hvordan den analytiske og den numeriske hessematrix kan kombineres, så styrkerne fra hver metode udnyttes bedst muligt.

For SAR-modellen $\mathbf{y} = \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ med loglikelihoodfunktionen

$$\begin{aligned} L &= -\frac{n}{2}\ln(2\pi\sigma^2) + \ln|I_n - \rho W| - \frac{e^T e}{2\sigma^2} \\ e &= \mathbf{y} - \rho W\mathbf{y} - X\boldsymbol{\beta} \\ \rho &\in (\lambda_{\min}^{-1}, 1) \end{aligned} \tag{5.4}$$

er hessematricen givet som i definition 5.9.

Definition 5.9 (Hessematrix for SAR-model):
Hessematricen H for SAR-modellen er givet ved

$$H = \begin{bmatrix} \frac{\partial^2 L}{\partial \rho^2} & \frac{\partial^2 L}{\partial \rho \partial \boldsymbol{\beta}^T} & \frac{\partial^2 L}{\partial \rho \partial \sigma^2} \\ \frac{\partial^2 L}{\partial \boldsymbol{\beta} \partial \rho} & \frac{\partial^2 L}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} & \frac{\partial^2 L}{\partial \boldsymbol{\beta} \partial \sigma^2} \\ \frac{\partial^2 L}{\partial \sigma^2 \partial \rho} & \frac{\partial^2 L}{\partial \sigma^2 \partial \boldsymbol{\beta}^T} & \frac{\partial^2 L}{\partial (\sigma^2)^2} \end{bmatrix},$$

hvor L er loglikelihood for SAR-modellen.

Sætning 5.10:
Den analytiske hessematrix $H^{(a)}$ for SAR-modellen er givet ved

$$H^{(a)} = \begin{bmatrix} -\operatorname{tr}(WAWA) - \frac{C}{\sigma^2} & -\frac{\mathbf{y}^T W^T X}{\sigma^2} & \frac{2C - B + 2\mathbf{y}^T W^T X \boldsymbol{\beta}}{2\sigma^4} \\ \cdot & -\frac{X^T X}{\sigma^2} & 0 \\ \cdot & \cdot & \frac{n}{2\sigma^4} - \frac{e^T e}{4\sigma^6} \end{bmatrix}, \tag{5.5}$$

hvor $A = (I_n - \rho W)^{-1}$, $B = \mathbf{y}^T(W + W^T)\mathbf{y}$ og $C = \mathbf{y}^T W^T W \mathbf{y}$. Prikkerne i (5.5) angiver, at matricen er symmetrisk, og at der på prikkernes indgange står det transponerede udtryk af de tilsvarende ingange på den anden side af diagonalen.

Bevis:
Vi ser kun på diagonalindgange, da udregninger for resten af matricen udføres på samme vis.

Første diagonalindgang i udtrykket fra definition 5.9 fås ved brug af (5.4) til

$$\frac{\partial^2 L}{\partial \rho^2} = \frac{\partial^2}{\partial \rho^2}\left(\ln\left|I_n - \rho W\right| - \frac{e^T e}{2\sigma^2}\right), \tag{5.6}$$

idet vi kun ser på de led, hvor ρ indgår. Vi ser på første led i (5.6), hvor (5.7) fremkommer ved anvendelse af regneregel C.1 i appendiks C, så

$$\begin{aligned} \frac{\partial}{\partial \rho}\left(\frac{\partial}{\partial \rho}\ln\left|I_n - \rho W\right|\right) &= \frac{\partial}{\partial \rho}\left(\left|I_n - \rho W\right|^{-1}\frac{\partial}{\partial \rho}\left|I_n - \rho W\right|\right) \\ &= \frac{\partial}{\partial \rho}\left(\left|I_n - \rho W\right|^{-1}\left|I_n - \rho W\right|\operatorname{tr}\left(-\left(I_n - \rho W\right)^{-1} W\right)\right) \\ &= \frac{\partial}{\partial \rho}\left(\operatorname{tr}\left(-\left(I_n - \rho W\right)^{-1} W\right)\right) \\ &= \operatorname{tr}\left(-\frac{\partial}{\partial \rho}\left(I_n - \rho W\right)^{-1} W\right). \end{aligned} \tag{5.7}$$

Nu anvendes regneregel C.2 i appendiks C, og det fås, at

$$\begin{aligned}
\operatorname{tr}\left(-\frac{\partial}{\partial\rho}\left(I_n-\rho W\right)^{-1}W\right) &= \operatorname{tr}\left(\left(I_n-\rho W\right)^{-1}\left(\frac{\partial}{\partial\rho}\left(I_n-\rho W\right)\right)\left(I_n-\rho W\right)^{-1}W\right)\\
&= \operatorname{tr}\left(-\left(I_n-\rho W\right)^{-1}W\left(I_n-\rho W\right)^{-1}W\right)\\
&= \operatorname{tr}\left(-W\left(I_n-\rho W\right)^{-1}W\left(I_n-\rho W\right)^{-1}\right) \qquad (5.8)\\
&= -\operatorname{tr}\left(W\left(I_n-\rho W\right)^{-1}W\left(I_n-\rho W\right)^{-1}\right),
\end{aligned}$$

hvor ligheden i (5.8) fremkommer ved anvendelse af regneregel C.3 i appendiks C.

Nu ses på sidste led i parentesen i (5.6), og idet vi kun ser på de led, hvor ρ indgår, fås det, at

$$\begin{aligned}
\frac{\partial^2 L}{\partial\rho^2}\left(-\frac{e^Te}{2\sigma^2}\right) &= \frac{\partial^2 L}{\partial\rho^2}\left(-\frac{\left(-\rho W\mathbf{y}\right)^T\left(-\rho W\mathbf{y}\right)}{2\sigma^2}\right)\\
&= \frac{\partial^2 L}{\partial\rho^2}\left(-\frac{\rho^2\left(W\mathbf{y}\right)^T\left(W\mathbf{y}\right)}{2\sigma^2}\right)\\
&= -\frac{\left(W\mathbf{y}\right)^T\left(W\mathbf{y}\right)}{\sigma^2}\\
&= -\frac{\mathbf{y}^TW^TW\mathbf{y}}{\sigma^2},
\end{aligned}$$

hvilket er lig med $-\frac{C}{\sigma^2}$. Dermed er udtrykket i den første diagonalindgang i hessematricen bevist.

Anden diagonalindgang i udtrykket fra definition 5.9 fås ved brug af (5.4) til

$$\frac{\partial^2 L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\left(-\frac{\left(\mathbf{y}-\rho W\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\mathbf{y}-\rho W\mathbf{y}-X\boldsymbol{\beta}\right)}{2\sigma^2}\right),$$

idet vi kun ser på de led, hvori $\boldsymbol{\beta}$ indgår. Det haves dermed, at

$$\frac{\partial^2 L}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T} = \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\left(-\frac{-\mathbf{y}^TX\boldsymbol{\beta}+\left(\rho W\mathbf{y}\right)^TX\boldsymbol{\beta}+\left(X\boldsymbol{\beta}\right)^TX\boldsymbol{\beta}-\left(X\boldsymbol{\beta}\right)^T\mathbf{y}+\left(X\boldsymbol{\beta}\right)^T\rho W\mathbf{y}}{2\sigma^2}\right). \qquad (5.9)$$

Kun et af leddene i (5.9) er ulig nul, når der differentieres to gange, nemlig

$$\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\left(-\frac{\left(X\boldsymbol{\beta}\right)^TX\boldsymbol{\beta}}{2\sigma^2}\right) = \frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\left(-\frac{\boldsymbol{\beta}^T\left(X^TX\right)\boldsymbol{\beta}}{2\sigma^2}\right).$$

Nu anvendes regneregel C.4 og C.5 i appendiks C, hvor vi sætter $A = X^TX$, så

$$\begin{aligned}
\frac{\partial^2}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}^T}\left(-\frac{\boldsymbol{\beta}^T\left(X^TX\right)\boldsymbol{\beta}}{2\sigma^2}\right) &= \frac{\partial}{\partial\boldsymbol{\beta}^T}\left(-\frac{2\boldsymbol{\beta}^T\left(X^TX\right)}{2\sigma^2}\right)\\
&= -\frac{X^TX}{\sigma^2}.
\end{aligned}$$

Dermed er udtrykket i den anden diagonalindgang i hessematricen bevist. Dermed er udtrykket i den anden diagonalindgang i hessematricen bevist.

Tredje diagonalindgang i hessematricen fra definition 5.9 er givet ved

$$\begin{aligned}\frac{\partial^2 L}{\partial(\sigma^2)^2} &= \frac{\partial^2}{\partial(\sigma^2)^2}\left(-\frac{n}{2}\ln(2\pi\sigma^2) - \frac{e^T e}{2\sigma^2}\right)\\ &= \frac{\partial}{\partial(\sigma^2)}\left(-\frac{2\pi n}{2}\cdot\frac{1}{2\pi\sigma^2} + \frac{2e^T e}{2\sigma^4}\right)\\ &= \frac{\partial}{\partial(\sigma^2)}\left(-\frac{n}{2\sigma^2} + \frac{e^T e}{\sigma^4}\right)\\ &= \frac{n}{2\sigma^4} - \frac{e^T e}{\sigma^6},\end{aligned}$$

hvor $e = \mathbf{y} - \rho W\mathbf{y} - X\boldsymbol{\beta}$.

Dermed er diagonalelementernes form i (5.5) bevist, og resten af elementerne i matricen udregnes på samme måde. ■

Hvis man har mange observationer n, så er det beregningsmæssigt vanskelige ved den analytiske hessematrix udregningen af det komplicerede udtryk $\operatorname{tr}(WAWA) = -\operatorname{tr}(W(I_n - \rho W)^{-1}W(I_n - \rho W)^{-1})$. I dette udtryk indgår, som det ses, den inverse af den tæt besatte $(n \times n)$-matrix $(I_n - \rho W)^{-1}$ og matrixmultiplikation med den n-dimensionale vægtmatrix W. Sådanne udregninger vil kræve $O(n^3)$ operationer at udføre, så derfor kan der anvendes numerisk integration på den første indgang i $H^{(a)}$. De resterende indgange i matricen $H^{(a)}$ er blot matrix-vektor-produkter med vægtmatricen som er tyndt besat, hvilket vil sige, at den kun indeholder få elementer forskellig fra nul.

Der er flere måde at gribe udregningen af det komplicerede udtryk an på. Vi ser på metoden med den kombinerede analytiske-numeriske beregning af hessematricen. Anvendelsen af $H^{(a)}$ og numerisk approksimation resulterer i en kombineret analytisk-numerisk Hessematrix $H^{(m)}$.

Idet vi benytter en univariat optimering af profilloglikelihoodfunktionen med hensyn til ρ, fås ikke den fulde numeriske hessematrix, men i stedet fås en numerisk hessematrix, der kun angår parameteren ρ, og som fremkommer af profilloglikelihoodfunktionen $L_p(\rho)$, så der er altså tale om hessematricen udtrykt ved $\frac{\partial^2 L_p}{\partial \rho^2}$. Det er dette udtryk, der repræsenterer estimatet af den andenordensafledede af profilloglikelihoodfunktionen. Profilloglikelihoodfunktionen kan i princippet godt anvendes til at finde korrekte værdier for parameteren ρ. Dog ønskes den fulde hessematrix H til den oprindelige loglikelihood L, der kan udtrykkes gennem parameteren ρ og vektoren $\boldsymbol{\theta} = \begin{bmatrix}\boldsymbol{\beta} & \sigma^2\end{bmatrix}^T$. Hessematricen H kan derfor reduceres til

$$H = \begin{bmatrix} \frac{\partial^2 L}{\partial \rho^2} & \frac{\partial^2 L}{\partial \rho \partial \boldsymbol{\theta}^T} \\ \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \rho} & \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}\cdot \end{bmatrix}.$$

Det er ifølge [LeSage and Pace, 2009, s. 58-59] muligt at tilpasse hessematricen for den empiriske profilloglikelihoodfunktion, så den giver et brugbart element til den fulde

hessematrix, således at

$$\frac{\partial^2 L}{\partial \rho^2} = \frac{\partial^2 L_p}{\partial \rho^2} + \frac{\partial^2 L}{\partial \rho \partial \boldsymbol{\theta}^T} \left(\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right)^{-1} \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \rho}. \tag{5.10}$$

Udtrykket i (5.10) viser sig at være nemt at udregne og kan indsættes i den fulde hessematrix, så det fås, at

$$H = \begin{bmatrix} \frac{\partial^2 L_p}{\partial \rho^2} + \frac{\partial^2 L}{\partial \rho \partial \boldsymbol{\theta}^T} \left(\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right)^{-1} \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \rho} & \frac{\partial^2 L}{\partial \rho \partial \boldsymbol{\theta}^T} \\ \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \rho} & \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}. \end{bmatrix}.$$

Vi har således udskiftet det komplicerede udtryk $H_{11}^{(a)}$ i den analytiske hessematrix med udtrykket for den tilpassede, empiriske profilloglikelihoodfunktion. Maksimum likelihood-estimeringen giver allerede en vektor med værdierne fra profilloglikelihood-funktionen som en funktion af parameteren ρ. Givet denne vektor af profilloglikelihood-funktioner, så er $\frac{\partial^2 L_p}{\partial \rho^2}$ meget nem at udregne.

Den kombinerede analytiske-numeriske hessematrix $H^{(m)}$ er dermed udtrykt ved

$$H^{(m)} = \begin{bmatrix} \frac{\partial^2 L_p}{\partial \rho^2} + \frac{\partial^2 L}{\partial \rho \partial \boldsymbol{\theta}^T} \left(\frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right)^{-1} \frac{\partial^2 L}{\partial \boldsymbol{\theta} \partial \rho} & -\frac{\mathbf{y}^T W^T X}{\sigma^2} & \frac{2C-B+2\mathbf{y}^T W^T X \boldsymbol{\beta}}{2\sigma^4} \\ . & -\frac{X^T X}{\sigma^2} & 0 \\ . & . & \frac{n}{2\sigma^4} - \frac{e^T e}{4\sigma^6} \end{bmatrix},$$

Den kombinerede analytiske-numeriske hessematrix kan anvendes til inferens for alle modellens parametre, idet varians-kovarians-matricen tilhørende parameterestimaterne som tidlligere nævnt kan antages at være lig med $-H^{-1}$, så givet $H^{(m)}$ kan parameterestimaterne simuleres ved anvendelse af multivariate normalfordelinger.

5.5 Residualer

Dette kapitel er baseret på [Olofsson, 2005, s. 372], [Encyclopedia, d] og [Draper and Smith, 1998, s. 60-67].

Når man har anvendt en model på et datasæt, er det relevant at teste, hvor god denne model er til at beskrive datasættet, for at se om der for eksempel skal tilføjes andre modelparametre eller om nogle modelparametre bør udelades. Sådanne modeltest kan laves ved hjælp af residualer, hvor de observerede data sammenlignes med dem modellen har forudsagt. Det er vigtigt at skelne mellem den ellers kendte afvigelse af data fra den sande, ikke-observerbare værdi og residualerne. Residualerne udtrykker forskellen mellem værdierne i datasættet og de af den anvendte model tilsvarende estimerede funktionsværdier. Residualer er hermed et udtryk for forskellen mellem data og modellens forudsigelse.

Definition 5.11:
Residualet **r** for SAR-modellen er givet ved

$$\mathbf{r} = \mathbf{y} - (I_n - \hat{\rho} W)^{-1} X \hat{\boldsymbol{\beta}},$$

hvor **y** er en datavektor, og $(I_n - \rho W)^{-1} X \hat{\boldsymbol{\beta}}$ er modellens middelværdi indeholdende estimaterne $\hat{\boldsymbol{\beta}}$ og $\hat{\rho}$.

Hvis data følger en anvendt model nøjagtigt, sådan at $\mathbf{Y} \sim N_n((I_n - \rho W)^{-1} X \hat{\boldsymbol{\beta}}, \sigma^2 I)$, så haves, at $\mathbf{r} \sim N_n(\mathbf{0}, \sigma^2 I)$. I praksis følger data ikke modellen fuldstændigt, men her vil residualer, der approksimativt er normalfordelte, understøtte valget af den anvendte model, men de kan ikke direkte bekræfte modellen. De viser blot, at vi ikke kan sige, at modellen er forkert på baggrund af datasættet. Der kan være problemer i forbindelse med at plotte de almindelige residualer, da residualer kan variere meget fra datapunkt til datapunkt. Derfor kan nogle af residualerne være meget store og andre små, hvilket kan gøre det problematisk at analysere, hvilke residualer der afviger specielt meget ud fra et plot af residualerne. Hermed laves ofte en skalering af residualerne, så de bliver standard normalfordelte. Disse kaldes for standardiserede residualer.

Definition 5.12:
De standardiserede residualer **R** for SAR-modellen er givet ved

$$\mathbf{R} = \frac{\mathbf{r}}{\sigma},$$

hvor **r** er de almindelige residualer.

Hvis modellen er god til at beskrive det pågældende datasæt, vil de standardiserede residualer være approksimativt uafhængige og standard normalfordelte, så $\mathbf{R} \sim N(0,1)$. Hermed fås et billede af residualerne, der ligger jævnt spredt uden særlige mønstre, hvilket vil sige, at de ligger spredt omkring en middelværdi med en konstant varians, og desuden må de ikke ligge mere en cirka tre standardafvigelser væk fra middelværdien.

Der findes forskellige, grafiske måder at plotte residualer på. Man kan blandt andet se på residualer i forhold til tid, estimeret data eller baggrundsvariable. Der undersøges, om der er *outliers*, der er afvigelser, som ligger tre eller flere standardafvigelser væk fra middelværdien for residualerne. Ofte undersøges outliers for at finde ud af, om de er udtryk for en tastefejl eller målefejl i datasættet, og som måske kan udelades, eller om de udtrykker, at man bør ændre den anvendte model. Ligeledes kan man undersøge *residualmønstre*, ved at plotte residualerne vertikalt med hensyn til datas orden efter tid, hvis denne er kendt, det tilsvarende estimat for middelværdien, ved brug af modellen eller den tilsvarende baggrundsvariabel. I alle tilfældene vil et tilfredsstillende plot repræsentere et horisontalt bånd af jævnt fordelte punkter. Et ikke tilfredsstillende plot kan se ud på forskellige måder, som hver repræsenterer forskellige fejl i residualerne.

Der er hovedsageligt tre typiske ikke-tilfredsstillende plots - båndet af residualer bliver bredere eller smallere, båndet af residualer bevæger sig opad eller nedad, eller båndet af residualer buer enten op eller ned. Ikke tilfredsstillende plots kan forårsages af forskellige fejlkilder, såsom et muligt behov for en transformation af Y, en ikke-konstant varians eller afhængighed i data, der ikke er taget højde for i modellen. Residualer kan dermed give et hint til, om modellen er forkert.

Modeltest med residualer varierer fra model til model og fra datasæt til datasæt, og vurderingen af residualerne, og hvad der gøres ved ikke-tilfredsstillende plots, afhænger af, hvad der vides eller formodes om det analyserede datasæt og den anvendte model.

Bayesianske, økonometriske, rumligt afhængige modeller 6

Dette kapitel er skrevet på baggrund af [LeSage and Pace, 2009, kapitel 5], [Lee, 2004, kapitel 2 og 3] og [Olofsson, 2005, afsnit 6.9].

Den bayesianske metode fokuserer på fordelinger, der indeholder information om både parametre og data ved et eksperiment, for at finde en posteriorfordeling, som kan nedbrydes til en mængde af marginale fordelinger. Disse betingede fordelinger karakteriserer fordelingen for en enkelt parameter givet de andre parametre i modellen. Markov Chain Monte Carlo-estimation (MCMC) forsimpler komplicerede estimationsproblemer til mere enkle problemstilinger, der baserer sig på de betingede fordelinger for hver parameter i en anvendt model. Således undgås en estimationsmetode, der bygger på en analytisk løsning af posteriorfordelingen indeholdende alle parametre i modellen.

6.1 Den bayesianske metode

I dette afsnit introduceres den bayesianske, økonometriske tilgangsvinkel til estimering af parametre i en rumlig model. Ved den bayesianske tilgangsvinkel er der både fokus på en fordeling for data såvel som for parametrene i den anvendte model, der beskriver data. Der tages udgangspunkt i Bayes formel, der kombinerer fordelingen for data, der er udtrykt ved en likelihoodfunktion, og priorfordelinger for modelparametrene, der udtrykker en forhåndsviden eller mangel på samme om disse. Likelihoodfunktionen og priorfordelingerne giver tilsammen en posteriorfordeling for modelparametrene givet data. Posteriorfordelingen er udgangspunktet for al inferens, idet den indeholder al relevant information med hensyn til estimationsproblemet. Relevant information inkluderer både information om data indeholdt i likelihoodfunktionen såvel som a priori-viden indeholdt i priorfordelingen for parametrene.

Den bayesianske tilgangsvinkel til estimering bygger på simpel sandsynlighedsregning.

Sætning 6.1 (Bayes sætning):
For to hændelser A og B er sandsynligheden for B betinget med A givet ved

$$p(B|A) = \frac{p(A|B)\,p(B)}{p(A)}. \tag{6.1}$$

Bevis:
For to stokastiske variable A og B kan den simultane sandsynlighed $p(A,B)$ udtrykkes ved hjælp af de betingede sandsynligheder $p(A|B)$ eller $p(B|A)$ og den marginale

sandsynlighed $p(B)$ eller $p(A)$, så

$$p(A,B) = p(A|B)\,p(B)$$
$$p(A,B) = p(B|A)\,p(A).$$

Hvis disse sættes lig med hinanden fås, at

$$p(A,B) = p(A|B)\,p(B) = p(B|A)\,p(A) \Rightarrow p(B|A) = \frac{p(A|B)\,p(B)}{p(A)}.$$

■

Vi lader $\mathcal{D} = \{\mathbf{y}, X, W\}$ repræsentere modeldata og $\boldsymbol{\theta} = \left[\theta_1, \theta_2, \ldots, \theta_n\right]^T$ repræsentere modelparametre, således at

$$p(\boldsymbol{\theta}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})}{p(\mathcal{D})}. \tag{6.2}$$

Det antages, at parameteren $\boldsymbol{\theta}$ har en priorfordeling $p(\boldsymbol{\theta})$, der afspejler den givne a priori-viden såvel som usikkerhed, der haves forud for observeret data, om parametrene i en model. Hvis der kun vides lidt fra tidligere eksperimenter eller kun haves lidt subjektiv viden om parametrene, så kan priorfordelingen udgøres af en vag fordeling som for eksempel en uniform fordeling, hvor alle udfald har lige stor sandsynlighed. Hvis der derimod haves stor a priori-viden, kan priorfordelingen være givet ved en mere detaljeret og beskrivende fordeling centreret om parameterværdier opnået fra tidligere eksperimenter. Venstresiden i (6.2) givet ved $p(\boldsymbol{\theta}|\mathcal{D})$ angiver posteriorfordelingen for $\boldsymbol{\theta}$ givet data og beskriver den opdaterede priorfordeling for parametrene $\boldsymbol{\theta}$ i modellen efter at have betinget med data fra det nye eksperiment. Det ses ved højresiden i (6.2), at posteriorfordeling er udtrykt ved et kompromis mellem priorfordelingen $p(\boldsymbol{\theta})$ og likelihoodfunktionen for dataene givet parametrene $p(\mathcal{D}|\boldsymbol{\theta})$. Den bayesianske tilgangsvinkel bygger på betingede sandsynligheder, hvilket gør os i stand til at opdatere vores a priori-viden om en ukendt størrelse $\boldsymbol{\theta}$ ved brug af model og data.

Hvis rumligt afhængigt data udviser stor variation over områder, og vi kun er i stand til at samle et datasæt indeholdende få af sådanne områder, så er posteriorfordelingen mere afhængig af a prior-viden givet ved $p(\boldsymbol{\theta})$ og tillægger dermed denne en større betydning end likelihoodfunktionen for det lille, observerede datasæt $p(\mathcal{D}|\boldsymbol{\theta})$. Omvendt, hvis vores a priori-viden er meget begrænset, og vi har et stort datasæt til rådighed, så vil posteriorfordeling tillægge likelihoodfunktionen, der beskriver information fra datasættet, en meget større betydning, idet likelihoodfunktionen vil indikere værdier for $\boldsymbol{\theta}$ fordelt tæt omkring en specifik værdi.

Den bayesianske statistiske metode er som den klassiske metode bygget op af de fire elementer modelspecifikation, modelestimering, modeltest og forecast. Specielt i forbindelse med rumlig økonometri ses på de tre problemstillinger estimation og inferens med hensyn til modelparametrene, modelspecifikation og -sammenligning og forecasting.

Estimering og inferens med hensyn til modelparametre

Al bayesiansk inferens med hensyn til $\boldsymbol{\theta}|\mathcal{D}$ er baseret på posteriorfordelingen $p(\boldsymbol{\theta}|\mathcal{D})$ for parameteren $\boldsymbol{\theta}$ givet data $\mathcal{D}$. Vi kan simplificere udtrykket i (6.2), idet fordelingen for data $p(\mathcal{D})$ kan opfattes som en konstant med hensyn til parameteren $\boldsymbol{\theta}$, så det haves, at

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto p(\mathcal{D}|\boldsymbol{\theta})\, p(\boldsymbol{\theta}).$$

Det vil sige, at posteriorfordelingen for $\boldsymbol{\theta}$ er proportionalt med produktet af likelihood-funktionen $p(\mathcal{D}|\boldsymbol{\theta})$ og priorfordelingen $p(\boldsymbol{\theta})$. Den samlede priorfordeling $p(\boldsymbol{\theta})$ for alle modelparametrene kan splittes op og udtrykkes som flere priorfordelinger $p(\theta_1)\, p(\theta_2)\ldots$, hvis modelparametrene er uafhængige.

Modelspecifikation og -sammenligning

Vi ser nu på, hvorledes der foretages modelsammenligning, og vi er her interesserede i at finde ud af, hvilken model der er mest sandsynlig givet data. Givet en mængde $i = 1,\ldots,m$ af bayesianske modeller, så vil hver af disse være repræsenteret af en likelihoodfunktion og priorfordeling, som

$$p\left(\boldsymbol{\theta}^i|\mathcal{D}\,, M_i\right) = \frac{p\left(\mathcal{D}|\boldsymbol{\theta}^i, M_i\right) p\left(\boldsymbol{\theta}^i|M_i\right)}{p(\mathcal{D}|M_i)},$$

hvor M_i udtrykker en modelspecifikation.

Ved at se på posteriorfordelingen betinget med modelspecifikationen M_i kan vi anvende Bayes formel og dermed til sidst udvide udtrykket $p(\mathcal{D}|M_i)$, som er den marginale likelihoodfunktion, på en måde tilsvarende (6.1). Vi er netop interessede i $p(\mathcal{D}|M_i)$, da den er nødvendig til modelsammenligning. Vi får nu en mængde af posterior modelsandsynligheder, der er ubetingede af $\boldsymbol{\theta}^i$, givet ved

$$p(M_i|\mathcal{D}) = \frac{p(\mathcal{D}|M_i)\, p(M_i)}{p(\mathcal{D})}, \tag{6.3}$$

der kan anvendes til modeltest med hensyn til forskellige modeller givet data. I (6.3) indgår de tre udtryk $p(\mathcal{D}|M_i)$,$p(M_i)$ og $p(\mathcal{D})$, hvoraf $p(\mathcal{D})$ i princippet blot er en konstant, og hyperpriorfordelingen $p(M_i)$ kan sættes til en uniform fordeling, således at der a priori antages lige stor sandsynlighed for alle modeller. Fordelingen $p(M_i)$ kaldes en hyperpriorfordeling for at skelne mellem de almindelige priorfordelinger for parametrene ρ, $\boldsymbol{\beta}$ og σ^2 i en anvendt model og priorfordelingerne for hyperparametrene indeholdt i selve priorfordelingen for ρ, $\boldsymbol{\beta}$ og σ^2, såsom middelværdi og varians i en normalfordeling eller a og b i den inverse gammafordeling (se appendiks B.4). Den marginale likelihoodfunktion $p(\mathcal{D}|M_i)$ findes ved integralet

$$p(\mathcal{D}|M_i) = \int_{\Theta^i} p\left(\mathcal{D}|\boldsymbol{\theta}^i, M_i\right) p\left(\boldsymbol{\theta}^i|M_i\right) d\boldsymbol{\theta}^i, \tag{6.4}$$

hvor $\boldsymbol{\theta}^i \in \Theta^i$, og Θ^i er mængden af mulige elementer for $\boldsymbol{\theta}$.

Maximum likelihood-metoder til modelsammeligning bygger på to likelihoodfunktioner for to modeller evalueret i maximum likelihood-estimatet for parametrene. Det betyder,

at modeltest med hensyn til modelsammeligning afhænger af skalarværdier af parameterestimater anvendt til at evaluere likelihoodfunktionen. I modsætning hertil anvender bayesiansk modelsamenligning hele posteriorfordelingen af værdier for parametrene. En ulempe ved bayesiansk modelsammenligning er, at der skal integreres over hele parametervektoren $\boldsymbol{\theta}$ i udtrykket (6.4).

Forecasting

Lad $\tilde{\mathbf{y}}$ være en observation simuleret ud fra observationer i det oprindelige datasæt $\mathcal{D}$. Posteriorfordelingen for forecasting givet datasættet $\mathcal{D}$ og modelparametre $\boldsymbol{\theta}$ er givet ved

$$p\left(\tilde{\mathbf{y}}|\mathcal{D}\right) = \int_{\Theta} p\left(\tilde{\mathbf{y}},\boldsymbol{\theta}|\mathcal{D}\right) d\boldsymbol{\theta} = \int_{\Theta} p\left(\tilde{\mathbf{y}}|\mathcal{D},\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}|\mathcal{D}\right) d\boldsymbol{\theta}.$$

Ovenstående er omskrevet ved at udtrykke den simultane fordeling ved en betinget fordeling og en marginal fordeling.

Når der haves store datasæt og ingen synderlig a priori-viden, så er der ikke stor forskel på inferens med hensyn til parametrene $\boldsymbol{\theta}$ foretaget ud fra den bayesianske metode og maximum likelihood-estimation, idet begge metoder i dette tilfælde næsten kun baseres på datasættet og dermed likelihoodfunktionen.

6.2 Bayesiansk fremgangsmåde til SAR-modellen

I dette afsnit introduceres den bayesianske metode til analytisk at estimere SAR-modellens parametre. Der ses på, hvordan inddragelsen af a priori-viden (som set i forhold til almindelige, statistiske tilgangsvinkler kan være en ulempe) spiller en mindre rolle i arbejdet med økonometriske, rumligt afhængige modeller, og dermed også når posteriorfordelingen skal findes, idet datasæt inden for økonomi ofte er så store, at likelihoodfunktionen dominerer. Vi illustrerer nu kombinationen af a priori-viden og datainformation ved at se på SAR-modellen.

Likelihoodfunktionen for SAR-modellen kendes fra sætning 3.6 og kan i den bayesianske sammenhæng noteres som i definition 6.2.

Definition 6.2:
Tæthedsfunktionen $p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)$ er defineret som likelihoodfunktionen fra SAR-modellen, så

$$p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |I_n - \rho W| \exp\left\{\frac{-1}{2\sigma^2}\left(\left(I_n-\rho W\right)\mathbf{y} - X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y} - X\boldsymbol{\beta}\right)\right\}.$$

Vi skal nu specificere en priorfordeling π, der afspejler vores a priori-viden, for parametrene i modellen, og priorfordelingen vil kombineret med likelihoodfunktionen definere posteriorfordelingen. Det kan her være en fordel at opsplitte priorfordelingen

$p(\boldsymbol{\theta})$, så $\pi\left(\boldsymbol{\beta},\sigma^2,\rho\right) = \pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right)$, idet der antages uafhængighed. Idet parameteren ρ spiller en vigtig rolle i modellen og ofte er af stor interesse i forhold til inferens, specificerer vi en uniform priorfordeling for denne parameter inden for det passende interval $\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right)$, hvor $\lambda_{\min}$ og $\lambda_{\max}$ repræsenterer den mindste og største egenværdi for vægtmatricen W. I praksis anvendes som før omtalt ofte intervallet [0,1). Lad π repræsentere en priorfordeling. Det haves defor, at

$$\begin{aligned} \rho &\sim U\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right) \Rightarrow \\ \pi\left(\rho\right) &= \frac{1}{\lambda_{\max}^{-1}-\lambda_{\min}^{-1}}, \end{aligned} \tag{6.5}$$

for $\lambda_{\min}^{-1} < \rho < \lambda_{\max}^{-1}$, og hvor U er en uniform fordeling (se appendiks B.3).

Vi kan for eksempel vælge en normal-invers gamma priorfordeling (NIG) for parametrene $\boldsymbol{\beta}$ og σ^2. Denne form for priorfordeling kan skrives som den normale priorfordeling for $\boldsymbol{\beta}$ (se appendiks B.1) betinget med σ^2 og den inverse gammafordeling for σ^2 (se appendiks B.4), så priorfordelingen er givet ved

$$\begin{aligned} \boldsymbol{\beta},\sigma^2 &\sim NIG(\mathbf{c},\Sigma,a,b) \Rightarrow \\ \pi\left(\boldsymbol{\beta},\sigma^2\right) &= \pi\left(\boldsymbol{\beta}|\sigma^2\right)\pi\left(\sigma^2\right) = N_k\left(\mathbf{c},\sigma^2\Sigma\right) IG(a,b), \end{aligned} \tag{6.6}$$

hvor $\mathbf{c}$ og $\sigma^2\Sigma$ angiver henholdvist middelværdivektoren og kovariansmatricen for normalfordelingen for $\boldsymbol{\beta}$, og parametrene a og b er to konstanter i den inverse gammafordeling. Det haves, at

$$\begin{aligned} \boldsymbol{\beta}|\sigma^2 &\sim N_k\left(\mathbf{c},\sigma^2\Sigma\right) \Rightarrow \\ \pi\left(\boldsymbol{\beta}|\sigma^2\right) &= \frac{1}{(2\pi)^{k/2}|\sigma^2\Sigma|^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta}-\mathbf{c})^T\Sigma^{-1}(\boldsymbol{\beta}-\mathbf{c})\right\}, \end{aligned} \tag{6.7}$$

og

$$\begin{aligned} \sigma^2 &\sim IG(a,b) \Rightarrow \\ \pi\left(\sigma^2\right) &= \frac{b^a}{\Gamma(a)}\left(\sigma^2\right)^{-(a+1)} \exp\left\{-\frac{b}{\sigma^2}\right\}, \end{aligned} \tag{6.8}$$

hvor $\Gamma(\cdot)$ repræsenterer gammafunktionen $\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt$. Hermed kan udtrykket i (6.6) ved direkte indsættelse af (6.7) og (6.8) skrives som

$$\begin{aligned} \boldsymbol{\beta},\sigma^2 &\sim N_k\left(\mathbf{c},\sigma^2\Sigma\right) IG(a,b) \Rightarrow \\ \pi\left(\boldsymbol{\beta},\sigma^2\right) &= \frac{b^a}{(2\pi)^{k/2}|\Sigma|^{1/2}\Gamma(a)}\left(\sigma^2\right)^{-(a+(k/2)+1)} \exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right\}, \end{aligned} \tag{6.9}$$

hvor $\sigma^2, a, b > 0$.

Sætning 6.3:
Når parametrene ρ, $\boldsymbol{\beta}$ og σ^2 har priorfordelingerne

$$\begin{aligned}\rho &\sim U\left(\lambda_{\min}^{-1}, \lambda_{\max}^{-1}\right)\\ \boldsymbol{\beta},\sigma^2 &\sim NIG\left(\mathbf{c}, \Sigma, a, b\right),\end{aligned}$$

så er posteriorfordelingen $p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right)$ givet ved

$$p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) \propto \left(\sigma^2\right)^{-(a^*+(k/2)+1)} \left|I_n-\rho W\right| \exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\left(\boldsymbol{\beta}-\mathbf{c}^*\right)+2b^*\right)\right\},$$

hvor

$$\begin{aligned}\Sigma^* &= \left(X^TX+\Sigma^{-1}\right)^{-1}\\ a^* &= a+\frac{n}{2}\\ b^* &= b+\frac{\mathbf{c}^T\Sigma^{-1}\mathbf{c}+\mathbf{y}^T\left(I_n-\rho W\right)^T\left(I_n-\rho W\right)\mathbf{y}-\left(\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\mathbf{c}^*}{2}\\ \mathbf{c}^* &= \Sigma^*\left(X^T\left(I_n-\rho W\right)\mathbf{y}+\Sigma^{-1}\mathbf{c}\right).\end{aligned}$$

Bevis:
Fra Bayes formel ved vi, at posteriorfordelingen for SAR-modellen er givet ved

$$\begin{aligned}p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) &= \frac{p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right)}{p\left(\mathcal{D}\right)}\\ &\propto p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right),\end{aligned}$$

idet parametrene a priori antages for værende uafhængige, og fordi $p\left(\mathcal{D}\right)$ ikke afhænger af parametrene. For at udregne posteriorfordelingen skal vi bruge et resultat, der siger, at

$$\begin{aligned}&\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)+\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\\ &\equiv \left(\boldsymbol{\beta}-\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\left(\boldsymbol{\beta}-\mathbf{c}^*\right)+2b^*,\end{aligned}$$

hvor

$$\begin{aligned}\Sigma^* &= \left(X^TX+\Sigma^{-1}\right)^{-1}\\ a^* &= a+\frac{n}{2}\\ b^* &= b+\frac{\mathbf{c}^T\Sigma^{-1}\mathbf{c}+\mathbf{y}^T\left(I_n-\rho W\right)^T\left(I_n-\rho W\right)\mathbf{y}-\left(\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\mathbf{c}^*}{2}\\ \mathbf{c}^* &= \Sigma^*\left(X^T\left(I_n-\rho W\right)\mathbf{y}+\Sigma^{-1}\mathbf{c}\right).\end{aligned} \tag{6.10}$$

Så kan vi udregne posteriorfordelingen ved hjælp af definition 6.2, (6.5) og (6.9), og det fås, at

$$
\begin{aligned}
p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) &\propto p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right) \\
&= \left(\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\left|I_n-\rho W\right|\exp\left\{-\frac{1}{2\sigma^2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\right\}\right)\cdot \\
&\quad \left(\frac{b^a}{(2\pi)^{k/2}|\Sigma|^{1/2}\Gamma(a)}\left(\sigma^2\right)^{-(a+(k/2)+1)}\exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}\right)^T\Sigma^{-1}\left(\boldsymbol{\beta}-\mathbf{c}\right)+2b\right)\right\}\right)\cdot \\
&\quad \pi\left(\rho\right) \\
&\propto \left(\sigma^2\right)^{-(a^*+(k/2)+1)}\left|I_n-\rho W\right|\exp\left\{-\frac{1}{2\sigma^2}\left(\left(\boldsymbol{\beta}-\mathbf{c}^*\right)^T\left(\Sigma^*\right)^{-1}\left(\boldsymbol{\beta}-\mathbf{c}^*\right)+2b^*\right)\right\},
\end{aligned}
$$

idet alle konstanter er udeladt inklusiv $\pi\left(\rho\right)$, da den er antaget at være en uniform fordeling og dermed konstant, og hvor Σ^*, a^*, c^* og b^* er givet som i (6.10). ■

Det skal her bemærkes, at priorfordelingen NIG ikke virker som en konjugeret priorfordeling, som det er tilfældet for ikke-rumlige modeller. En konjugeret priorfordeling resulterer i en beregningsmæssigt håndterbar posteriorfordeling, hvor posteriorfordelingen antager samme form som priorfordelingen med den eneste ændring, at modelparametrene er opdaterede. Hvis det skulle gælde for SAR-modellen, skulle $\rho = 0$, så $I_n - \rho W = I_n$.

De parametre, der specificerer vores a priori-viden, er $\mathbf{c}$, Σ, a og b fra NIG. For ρ har vi angivet en priorfordeling, der med lige stor sandsynlighed tildeler ρ en værdi i intervallet $\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right)$. Vores a priori-viden kunne også have afspejlet forhåndsviden om ρ og dermed have indeholdt priorparametre for ρ, hvis vi havde valgt en mere informativ priorfordeling for denne modelparameter. For eksempel kunne anvendes en betafordeling, der er skaleret til intervallet $\left(\lambda_{\min}^{-1},\lambda_{\max}^{-1}\right)$ (se appendiks B.5). I de tilfælde, hvor vi har meget lidt a priori-viden, og der dermed hersker stor usikkerhed om parameteren $\boldsymbol{\beta}$, kan vi konstruere en priorfordelingen for $\boldsymbol{\beta}$, der ikke er informativ, ved at sætte $\mathbf{c} = \mathbf{0}$ og tildele parameteren $\boldsymbol{\beta}$ en meget stor varians og ingen kovarians mellem de enkelte parametre i vektoren $\boldsymbol{\beta}$ ved for eksempel at sætte $\Sigma = I_k \cdot 10^{10}$. Der er således tale om en uegentlig priorfordeling, hvilket vil sige, at der ikke er tale om en veldefineret tæthed, der summerer til 1, hvilket der skal tages højde for i beregningerne, således at der i sidste ende fås en veldefineret posteriorfordeling, der summerer til 1. Hvis vi ligeledes vil konstruere en priorfordeling for σ^2, der ikke er informativ, kan vi sætte $a,b \to 0$. Desuden antages uafhængighed mellem priorfordelingen for parameteren ρ og den for $\boldsymbol{\beta}$, σ^2. Det er vigtigt at bemærke, at uafhængighed mellem priorfordelingerne ikke nødvendigvis resulterer i uafhængighed i posteriorfordelingerne for modelparametrene.

Idet der i tilfældet med SAR-modellen er anvendt informative priorfordelinger, kompliceres tilfældet på to måder. For det første skal der specificeres eller tildeles værdier for parametrene i NIG priorfordelingen. For det andet er posteriorfordelingen i sætning 6.3 svær at analysere og fortolke, fordi den kræver integration med hensyn til σ^2 og ρ for at få posteriorfordelingen for $\boldsymbol{\beta}$, og tilsvarende skal der integreres med hensyn til $\boldsymbol{\beta}$ og ρ for at

finde posteriorfordelingen for σ^2 osv. Man kan se på, hvorvidt alle priorfordelingerne har nogen indflydelse på inferensen med hensyn til $\boldsymbol{\beta}$, og hvis ikke kan posteriorfordelingen simplificeres ved at indføre priorfordelinger, der ikke er informative, i udtrykket i sætning 6.3.

Hvis vi kendte middelværdierne for posteriorfordelingerne for ρ og $\mathbf{c}^*$, ville det sidste led i b^* være lig med den kvadrerede sum af residualerne fra SAR-modellen. Det indikerer, at forholdet $\frac{b^*}{a^*}$ approksimativt er lig med den kvadrerede sum af residualerne, når $a,b,\Sigma^{-1} \to 0$. Disse tiltag for NIG priorfordelingerne ville resultere i, at parametrene $\boldsymbol{\beta}$ og σ^2 ville blive tildelt priorfordelinger, der ikke er informative.

Bayesiansk løsning til estimering af parameteren ρ i SAR-modellen

Vi vil nu finde et estimat for ρ i SAR-modellen, og vi erstatter NIG priorfordelingen fra før med en priorfordeling, der ikke er informativ og dermed repræsenterer en fuldstændig usikkerhed (se appendiks B.5). Det gøres ved at lade hyperparametrene gå mod gammafunktionens grænseværdi nul, så $a, b \to 0$, hvilket gør, at gammafordelingen bliver flad, og der haves en uegentlig tæthed, mens kovariansmatricen for normalfordelingen $\Sigma^{-1} \to$ nulmatricen. Der antages desuden uafhængighed mellem priorfordelingen for ρ og priorfordelingen for $\boldsymbol{\beta}$ og σ^2, så $\pi\left(\boldsymbol{\beta},\sigma^2,\rho\right) = \pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right)$. Ved at indsætte dette i sætning 6.3 fås posteriorfordelingen til

$$\begin{aligned} p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right) &\propto p\left(\mathcal{D}|\boldsymbol{\beta},\sigma^2,\rho\right)\pi\left(\boldsymbol{\beta},\sigma^2\right)\pi\left(\rho\right) \\ &\propto \sigma^{-n-1}\left|I_n-\rho W\right|\exp\left\{-\frac{1}{2\sigma^2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\right\}\pi\left(\rho\right). \end{aligned} \tag{6.11}$$

Vi kan opfatte σ^2 som en sekundær parameter, der ikke har den primære interesse for nu, og integrere den ud af udtrykket i (6.11) ved hjælp af egenskaber for den inverse gammafordeling. Det gøres ved at udnytte egenskaber for den inverse gammafordeling, så

$$\begin{aligned} p\left(\boldsymbol{\beta},\rho|\mathcal{D}\right) &\propto \left|I_n-\rho W\right|\left(\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\right)^{-n/2}\pi\left(\rho\right) \\ &\propto \left|\left(I_n-\rho W\right)\right|\left(\left(n-k\right)s^2+\left(\boldsymbol{\beta}-\mathbf{c}^{**}\right)^t X^T X\left(\boldsymbol{\beta}-\mathbf{c}^{**}\right)\right)^{-n/2}\pi\left(\rho\right), \end{aligned} \tag{6.12}$$

hvor

$$\begin{aligned} \mathbf{c}^{**} &= \left(X^T X\right)^{-1} X^T\left(I_n-\rho W\right)\mathbf{y} \\ s^2 &= \frac{\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\mathbf{c}^{**}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\mathbf{c}^{**}\right)}{n-k}. \end{aligned}$$

Hvis der betinges med ρ, så repræsenterer udtrykket i (6.12) en multivariat t-fordeling, som vi kan integrere med hensyn til $\boldsymbol{\beta}$, idet denne er en tæthedsfunktion, der integrerer til 1. Således fås den marginale posteriorfordeling for ρ, hvor

$$p\left(\rho|\mathcal{D}\right) \propto \left|I_n-\rho W\right|\left(s^2\right)^{-(n-k)/2}\pi\left(\rho\right). \tag{6.13}$$

Der findes ingen analytisk løsning til at finde middelværdien eller variansen for posteriorfordelingen for ρ, hvor vi er interesserede i begge dele i forbindelse med inferens. Dog kan der anvendes univariate, numeriske integrationsmetoder, således at vi kan finde både middelværdien og variansen for posteriorfordelingen for ρ. De integraler, der er tale om, er

$$\mathrm{E}\left[\rho|\mathcal{D}\right] = \int_{\Theta^i} \rho \cdot p\left(\rho|\mathcal{D}\right) d\rho \tag{6.14}$$

$$\mathrm{Var}\left[\rho|\mathcal{D}\right] = \int_{\Theta^i} \left(\rho - \mathrm{E}\left[\rho|\mathcal{D}\right]\right)^2 \cdot p\left(\rho|\mathcal{D}\right) d\rho, \tag{6.15}$$

hvor $\rho \in \Theta^i$. Hvis udtrykket i (6.13) sættes direkte ind på pladsen for sandsynligheden i ovenstående udtryk, skal der divideres med integralet $\int_\Theta p\left(\rho|\mathcal{D}\right) d\rho$ (det, der integreres til) for at få en normeret tæthed, der integrerer til 1, så

$$\mathrm{E}\left[\rho|\mathcal{D}\right] = \frac{\int_\Theta \rho \cdot p\left(\rho|\mathcal{D}\right) d\rho}{\int_\Theta p\left(\rho|\mathcal{D}\right) d\rho}$$

$$\mathrm{Var}\left[\rho|\mathcal{D}\right] = \frac{\int_\Theta \left(\rho - \mathrm{E}\left[\rho|\mathcal{D}\right]\right)^2 \cdot p\left(\rho|\mathcal{D}\right) d\rho}{\int_\Theta p\left(\rho|\mathcal{D}\right) d\rho}.$$

Det ses ved (6.13), (6.14) og (6.15), at integrationen involverer udregning af determinanten af $(n \times n)$-matricen $I_n - \rho W$ over et interval med passende værdier for ρ. Her kan der anvendes Monte Carlo-approksimation af logdeterminanten ved brug af (4.5) i kapitel 4. Der skal desuden vælges et passende interval for ρ. Hvis der ikke haves a priori-viden, kan der vælges et interval med den mindste og største egenværdi for vægtmatricen W som grænser, som angiver de teoretiske passende værdier for ρ. For store datasæt er den rækkestokastiske vægtmatrix W tyndt besat og indeholder dermed en masse nuller, og da vi ved, at $\lambda_{\max} = 1$, er det kun $\lambda_{\min}$, der skal udregnes, hvis ikke denne blot sættes lig nul. Hvis der haves en vis a priori-viden, kan der anvendes en priorfordeling, der er begrænset til intervallet $(-1,1)$, hvilket afspejler en viden om, at mange tidligere undersøgelser har resulteret i estimater af ρ, der ligger inden for dette interval, og en viden om at hvis ρ befinder sig uden for dette interval, kan det være, at der er problemer med vægtmatricen eller den valgte model. Et eksempel på en priorfordeling, der kan anvendes til at begrænse ρ, er en skalleret betafordeling $B(d,d)$ (se appendiks B.5), så den er defineret på intervallet $(-1,1)$ og har middelværdien 0. Man kan også argumentere for at negative værdier for ρ, og dermed negativ korrelation mellem observationer, har lav interesse i et specifikt tilfælde, og der derfor kan vælges en priorfordeling, der begrænser ρ til intervallet $[0,1)$ som for eksempel en skaleret betafordeling. Således undgås det at udregne den mindste egenværdi for den store vægtmatrix W.

I stedet for at se på den oprindelige priorfordeling for ρ i (6.13) kan vi se på ln for dette udtryk. Der indføres en $(q \times 1)$-vektor af værdier fra $[\rho_{\min},\rho_{\max}]$, så $\boldsymbol{\rho} = [\rho_1 \quad \dots \quad \rho_q]$ indeholder tilfældigt udvalgte værdier for ρ, og således fås

$$\begin{bmatrix} \ln p\left(\rho_1|\mathbf{y}\right) \\ \ln p\left(\rho_2|\mathbf{y}\right) \\ \vdots \\ \ln p\left(\rho_q|\mathbf{y}\right) \end{bmatrix} = \kappa + \begin{bmatrix} \ln\left|I_n - \rho_1 W\right| \\ \ln\left|I_n - \rho_2 W\right| \\ \vdots \\ \ln\left|I_n - \rho_q W\right| \end{bmatrix} - \frac{n-k}{2} \begin{bmatrix} \ln\left(s^2\left(\rho_1\right)\right) \\ \ln\left(s^2\left(\rho_2\right)\right) \\ \vdots \\ \ln\left(s^2\left(\rho_q\right)\right) \end{bmatrix},$$

hvor

$$\begin{aligned} s^2(\rho) &= e_0^T e_0 - 2\rho_i e_d^T e_0 + \rho_i^2 e_d^T e_d \\ e_0 &= \mathbf{y} - Xc_0 \\ e_d &= W\mathbf{y} - Xc_d \\ c_0 &= \left(X^T X\right)^{-1} X^T \mathbf{y} \\ c_d &= \left(X^T X\right)^{-1} X^T W\mathbf{y}. \end{aligned}$$

Ved brug af denne vektor kan nu anvendes univariat numerisk integration ved brug af for eksempel Simpsons metode. Dog skal vi stadig integrere to gange for at få både middelværdi og varians for posterioren for ρ.

For eksempel kræves der ikke numerisk integration til at finde middelværdien for posteriorfordelingen for $\boldsymbol{\beta}$, der ved mindste kvadraters metode er givet ved $\mathrm{E}\left[\boldsymbol{\beta}|\mathbf{y},X,W\right] = \mathbf{c}^{**} = \left(X^T X\right)^{-1} X^T \left(I_n - \mathrm{E}\left[\rho|\mathcal{D}\right] W\right)\mathbf{y}$. Hvis der ønskes en varians for posteriorfordelingen for $\boldsymbol{\beta}$, skal der udføres numerisk integration med henblik på denne. Det samme gør sig gældende for parameteren σ^2.

6.3 MCMC anvendt på bayesiansk SAR-model

I dette afsnit ses på Markov Chain Monte Carlo-estimation for SAR-modellen. Man anvender ofte MCMC, når der ikke haves konjugerede fordelinger og ved fordelinger, der er komplicerede at udføre beregning på. Ved den analytiske fremgangsmåde beskrevet ovenfor står og falder alt med, hvorvidt det er muligt at løse integraler, men ved anvendelse af MCMC, undgås den analytiske inddragelse af disse. Som tidligere nævnt kræver posteriorfordelingen for SAR-modellen univariat, numerisk integration for at få en middelværdi og varians for posteriorfordelingen for for eksempel parameteren ρ. MCMC-estimation er et alternativ til denne tilgangvinkel. Et generelt resultat jævnfør [LeSage and Pace, 2009, s. 123] er, at MCMC-sampling fra en mængde af betingede fordelinger for alle parametre i en model sammen med Gibbs-sampling resulterer i en mængde af estimater, der konvergerer mod den simultane posteriorfordeling for alle parametrene. Hvis posteriorfordelingen kan nebrydes til en mængde af betingede fordelinger for hver parameter i modellen, så kan vi simulere fra disse og få tilfredsstillende, bayesianske parameterestimater. MCMC-estimation bygger hermed på antagelsen om, at i stedet for at se på posteriortætheden for alle parametrene, kan vi se på store simuleringer fra stokastiske posteriorfordelinger.

Vi ser først på MCMC anvendt med Gibbs-sampling, der er en algoritme, som anvendes, når der er tale om pæne priorfordelinger, idet Gibbs-sampling tager udgangspunkt i betingede priorfordelinger for alle parametre i en model, således at der haves en priorfordeling for hver parameter, der er betinget med alle de resterende parametre i modellen. Gibbs-sampling anvendes ofte til estimering af parametrene $\boldsymbol{\beta}$ og σ^2, idet disse tit har en kendt priorfordeling. Gibbs-sampling giver en mængde af stokastiske simuleringer fra en multivariat fordeling, når direkte simulering er kompliceret. Derefter ses på MCMC anvendt med Metropolis-Hastings-sampling (M-H-sampling), som ofte

anvendes, når der ikke haves viden om, hvilken type fordeling en pågældende parameter har. M-H-sampling anvendes således ofte til estimering af parameteren ρ, idet det kan være svært at finde en kendt fordeling for denne parameter, men M-H-sampling kan i princippet anvendes i de fleste tilfælde. Hvis alle betingede priorfordelinger for alle parametre er håndterbare og kendte fordelinger, så anvendes en standard Gibbs-sampler, men hvis ikke så anvendes Metroplis-Hastings i Gibbs, således at M-H-sampling indgår som et led i Gibbs-sampling.

Simulering af fordelinger for parametrene $\boldsymbol{\beta}$ og σ^2 med Gibbs-sampling

Ud fra den simultane fordeling $p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right)$ i sætning 6.3 kan den betingede fordeling for hver af de tre parametre findes ved at evaluere $p\left(\boldsymbol{\beta},\sigma^2,\rho|\mathcal{D}\right)$ for hver enkelt parameter, mens de resterende parametre antages at være kendte og dermed holdes fast.

Nu ses på de betingede fordelinger for parametrene $\boldsymbol{\beta}$ og σ^2 i SAR-modellen baseret på NIG priorfordelingen for parameterene $\boldsymbol{\beta}$ og σ^2. Hvis ρ er kendt, så leder den konjugerede NIG priorfordeling for $\boldsymbol{\beta}$ og σ^2 til en simultan NIG posteriorfordeling (betinget af ρ) for $\boldsymbol{\beta}$ og σ^2. Når den simultane NIG posteriorfordeling kendes, haves også den k-dimensionale, betingede normale posteriorfordeling for $\boldsymbol{\beta}$ givet ved $N_k(\mathbf{c}^*,\Sigma^*)$ og den betingede posteriorfordeling for σ^2 givet ved $IG(a^*,b^*)$. I tilfældet, hvor alle parametre er ukendte, mangler vi en betinget fordeling for ρ, men for nu antager vi, at ρ er kendt og holdes fast. Tidligere er det blevet vist, hvordan ρ estimeres, når den er ukendt.

Nu ser vi på Gibbs-sampling, som er beskrevet ved følgende algoritme, og vi ser kun på parametrene $\boldsymbol{\beta}$ og σ^2, som skal estimeres. Vi har parametervektoren $\boldsymbol{\theta} = \left(\boldsymbol{\beta}_{(0)},\sigma^2_{(0)}\right)$, hvor indekset 0 angiver, at der er tale om begyndelsesbetingelser for parametrene. Givet en begyndelsesbetingelse for σ^2 givet ved $\sigma^2_{(0)}$ og med ρ givet kan middelværdien $\mathbf{c}^*$ og kovaransmatricen Σ^* udregnes for den betingede multivariate normalfordeling for $\boldsymbol{\beta}$ ved at anvende udtrykket

$$p\left(\boldsymbol{\beta}|\rho,\sigma^2_{(0)}\right) \sim N_k\left(\mathbf{c}^*,\sigma^2_{(0)}\Sigma^*\right), \tag{6.16}$$

hvor

$$\begin{aligned}\Sigma^* &= \left(X^TX+\Sigma^{-1}\right)^{-1}\\ \mathbf{c}^* &= \Sigma^*\left(X^T\left(I_n-\rho W\right)\mathbf{y}+\Sigma^{-1}\mathbf{c}\right).\end{aligned}$$

Ved hjælp af en algoritme, der producerer vektorer af stokastiske multivariate normale vektorer med middelværdi og varians-kovariansmatrix som i (6.16), fås den simulerede værdi $\boldsymbol{\beta}_{(1)}$, som vi kan erstatte begyndelsesværdien $\boldsymbol{\beta}_{(0)}$ med.

Det samme gør sig gældende for parameteren σ^2. Givet ρ og den simulerede værdi for $\boldsymbol{\beta}$ givet ved $\boldsymbol{\beta}_{(1)}$ kan vi simulere værdierne a^* og b^* for den betingede, inverse gammafordeling for σ^2 ved at anvende udtrykket

$$p\left(\sigma^2|\rho,\boldsymbol{\beta}_{(1)}\sigma^2_{(0)}\right) \sim IG\left(a^*,b^{**}\right),$$

hvor

$$a^* = a + \frac{n}{2}$$
$$b^{**} = b + \frac{\left((I_n - \rho W)\mathbf{y} - X\boldsymbol{\beta}_{(1)}\right)^T \left((I_n - \rho W)\mathbf{y} - X\boldsymbol{\beta}_{(1)}\right)}{2}.$$

Algoritmen anvendes nu til at lave en stokastisk værdi fra fordelingen $IG(a^*, b^{**})$ for at opdatere begyndelsesværdien $\sigma^2_{(0)}$ til $\sigma^2_{(1)}$.

Nu vendes tilbage til det første skridt i algoritmen, hvor der anvendes den betingede fordeling for $\boldsymbol{\beta}$, således at der først findes endnu en værdi for $\boldsymbol{\beta}$ givet ved $\boldsymbol{\beta}_{(2)}$, som indsættes i udtrykket for σ^2, hvorved vi får endnu en værdi for σ^2 givet ved $\sigma^2_{(2)}$. Således fortsættes, indtil der haves en stor mængde af simulerede værdier for parametrene $\boldsymbol{\beta}$ og σ^2.

Mængden af de simulerede værdier taget fra deres respektive, betingede fordelinger repræsenterer ved ovenstående metode, hvor der anvendes den fulde mængde af alle betingede fordelinger for alle parametre i modellen, simuleringer fra den simultane posteriorfordeling for modelparametrene. Det skal gøres klart, at det er blevet bevist, at estimater for $\boldsymbol{\beta}$ og σ^2 fra en Gibbs-sampling, der har kørt længe nok, er de samme som ved udtagning af et estimat for hver parameter taget fra hele posteriorfordelingen ved hjælp af for eksempel M-H.

Efter ovenstående proces kan der laves inferens med hensyn til parametrene ved at se på middelværdien og variansen for mængden af alle de simulerede værdier.

Simulering af fordeling for ρ med M-H-sampling

Afsnittet er skrevet på baggrund af [Lee, 2004, s. 268-270].

Lad $p(\boldsymbol{\theta}|\mathcal{D})$ repræsentere posteriorfordelingen, hvor $\boldsymbol{\theta}$ er modelparametre, og $\mathcal{D}$ er data. Hvis der haves et tilstrækkeligt antal simuleringer fra $p(\boldsymbol{\theta}|\mathcal{D})$, kan formen på sandsynlighedstætheden approksimeres, således at det ikke er nødvendigt at finde den nøjagtige analytiske-numeriske form på sandsynlighedstætheden. Vi vælger en tæthed $f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^t\right)$ kaldet forslagsfordelingen for parametervektoren $\boldsymbol{\theta}^t$, der kan anvendes til at finde fordelingen for $\boldsymbol{\theta}^{t+1}$, således at fordelingen for $\boldsymbol{\theta}^{t+1}$ er fremkommet udelukkende ved kendskab til fordelingen for $\boldsymbol{\theta}^t$. Fordelingen $f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{t+1}\right)$ for $\boldsymbol{\theta}^{t+1}$ kan dernæst anvendes til at finde fordelingen for $\boldsymbol{\theta}^{t+2}$ og så fremdeles, hvilket vil sige, at der haves en markovkæde med en givet begyndelsesbetingelse $\boldsymbol{\theta}^0$. Den mest anvendte metode i forbindelse med MCMC er M-H-sampling. M-H-algoritmen starter med at simulere værdier fra den valgte tæthed $f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^t\right)$ til tiden $t+1$, givet at markovkæden er i $\boldsymbol{\theta}^t$, og det antages desuden, at

$$\int_{\boldsymbol{\theta}} f(\boldsymbol{\theta}|\boldsymbol{\theta}^t) = 1.$$

Der kan nu simuleres en kandidat $\boldsymbol{\theta}^*$ til punktet $\boldsymbol{\theta}^{t+1}$ fra tætheden $f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^t\right)$, hvor punktet $\boldsymbol{\theta}^*$ enten kan accepteres som $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^*$ eller forkastes, så vi sætter $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$. Om punktet

forkastes eller accepteres kommer an på tætheden $f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$. Hvis tætheden er reversibel, hvilket vil sige, hvis

$$p(\boldsymbol{\theta}^t|\mathcal{D})f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t) = p(\boldsymbol{\theta}^*|\mathcal{D})f(\boldsymbol{\theta}^t|\boldsymbol{\theta}^*),$$

så bliver den til den endelige tæthed, og sandsynligheden, for at $\boldsymbol{\theta}^*$ accepteres, er dermed den samme, som sandsynligheden for den forkastes. Hvis der derimod gælder, at

$$p(\boldsymbol{\theta}^t|\mathcal{D})f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t) > p(\boldsymbol{\theta}^*|\mathcal{D})f(\boldsymbol{\theta}^t|\boldsymbol{\theta}^*),$$

svarer det til, at der er større sandsynlighed for, at $\boldsymbol{\theta}^t$ forkastes til fordel for $\boldsymbol{\theta}^*$ end omvendt, hvilket resulterer i et ujævnt traceplot for $\boldsymbol{\theta}$, og det kan dermed være vanskeligt at vurdere, om der er konvergens. Det ønskes ofte at have en fordeling for $\boldsymbol{\theta}$, der bevæger sig jævnt på et traceplot, hvilket vil sige, at omkring halvdelen af de simulerede værdier accepteres. For at sørge for at der er lige stor sandsynlighed for, at $\boldsymbol{\theta}^t$ vælges frem for $\boldsymbol{\theta}^*$, som at $\boldsymbol{\theta}^*$ erstatter $\boldsymbol{\theta}^t$, indføres en sandsynlighed $\psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$, der opfylder, at

$$\begin{aligned} p(\boldsymbol{\theta}^t|\mathcal{D})f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)\psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t) &= p(\boldsymbol{\theta}^*|\mathcal{D})f(\boldsymbol{\theta}^t|\boldsymbol{\theta}^*) \Rightarrow \\ \psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t) &= \frac{p(\boldsymbol{\theta}^*|\mathcal{D})f(\boldsymbol{\theta}^t|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^t|\mathcal{D})f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)}. \end{aligned}$$

Da der ønskes en så høj acceptsandsynlighed som muligt, sættes $\psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$ til

$$\psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t) = \min\left[\frac{p(\boldsymbol{\theta}^*|\mathcal{D})f(\boldsymbol{\theta}^t|\boldsymbol{\theta}^*)}{p(\boldsymbol{\theta}^t|\mathcal{D})f(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)},1\right]. \tag{6.17}$$

Det vil sige, at $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^*$ med sandsynligheden (6.17) og ellers sættes $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$, hvilket er ensbetydende med, at kæden bliver stående i den nuværende værdi for $\boldsymbol{\theta}$.

M-H-algoritmen kan formelt skrives som gentagelser af følgende trin:

1. Antag kæden er i $\boldsymbol{\theta}^t$ (begyndelsesværdien for første gennemløb af algoritmen er et på forhånd valgt $\boldsymbol{\theta}^0$).
2. Simuler punktet $\boldsymbol{\theta}^*$ fra $p(\boldsymbol{\theta}|\mathcal{D})$.
3. Udregn $\psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$ fra (6.17).
4. Generer en værdi u fra en uniform fordeling $U(0,1)$.
5. Hvis $u \leq \psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$ sættes $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^*$ og hvis $u \geq \psi_H(\boldsymbol{\theta}^*|\boldsymbol{\theta}^t)$ sættes $\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t$.

Algoritmen giver en følge $\{\boldsymbol{\theta}^0,\boldsymbol{\theta}^1,\ldots\}$, der har længde efter antallet af gennemløb af algoritmen. De simulerede værdier, der fås inden følgen konvergerer, sorteres fra som burn-in. Burn-in er den tid, det tager algoritmen at finde en ligevægtsfordeling, hvilket vil sige, at resten af de simulerede værdier for parametrene efter burn-in approksimeret er konvergeret mod posteriorfordelingen.

Denne metode til simulering repræsenterer en Markovkæde, og jævnfør [LeSage and Pace, 2009, s. 123] konvergerer den mod den rigtige fordeling, og at den er i stand til at producere simuleringer fra posteriorfordelingen $p(\boldsymbol{\theta}|\mathcal{D})$. Vi kan derfor anvende M-H til at simulere fra betingede fordelinger, hvor fordelingens form ikke er kendt, og det

er præcis denne situation, vi er i med hensyn til parameteren ρ ved anvendelse af SAR-modellen. I andre situationer som for eksempel gør sig gældende for parametrene $\boldsymbol{\beta}$ og σ^2 i SAR-modellen, tager de betingede fordelinger form som for eksempel multivariate normalfordelinger med middelværdi og varians, der let kan beregnes ved hjælp af almindelig, lineær regression. Hvis fordelingerne er kendte, anvendes Gibbs-sampling, hvor der som før nævnt ses på betingede fordelinger.

Der findes forskellige metoder til at vælge forslagsfordelingen $f\left(\boldsymbol{\theta}|\boldsymbol{\theta}^0\right)$, hvilket kan ses i [Lee, 2004, s. 270].

Simulering af ρ

Indtil nu har det været antaget, at parameteren ρ har været kendt, men da det ikke er tilfældet i praksis, er det nødvendigt at estimere værdien for ρ. For SAR-modellen simuleres ρ fra sin betingede fordeling givet ved

$$\begin{aligned} p\left(\rho|\boldsymbol{\beta},\sigma^2\right) &\propto \frac{p\left(\rho,\boldsymbol{\beta},\sigma^2|\mathcal{D}\right)}{p\left(\boldsymbol{\beta},\sigma^2|\mathcal{D}\right)} \\ &\propto \left|I_n-\rho W\right|\exp\left\{-\frac{1}{2\sigma^2}\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)^T\left(\left(I_n-\rho W\right)\mathbf{y}-X\boldsymbol{\beta}\right)\right\}, \end{aligned} \tag{6.18}$$

hvor tæller og nævner er givet ved sætning 6.3 og (6.9), og $\mathcal{D}$ er mængden af data, der ønskes en model for. Idet den betingede fordeling ikke antager en kendt form, som det var tilfældet med de betingede fordelinger for $\boldsymbol{\beta}$ og σ^2, anvendes nu M-H-sampling for at finde ρ. Først antages en mulig fordeling, som ifølge [Lee, 2004, s. 270] for eksempel kan være en standard normalfordeling $N(0,1)$ sammen med en justeret random walk, fra hvilken der genereres en værdi for ρ noteret ρ^* givet ved

$$\rho^* = \rho^t + c\cdot N(0,1), \tag{6.19}$$

hvor ρ^t er ρ ved skridt t eller til tiden t, og konstanten c kaldes en justeringsparameter, der kan tilpasses, så M-H-sampling dækker hele den betingede fordeling. I praksis afprøves forskellige værdier af c for at se hvilken værdi, der virker bedst, hvilket uddybes i eksempel 6.4. Værdien ρ^* og værdien til tiden t, ρ^t, evalueres i (6.18) for at få en acceptsandsynlighed ved brug af

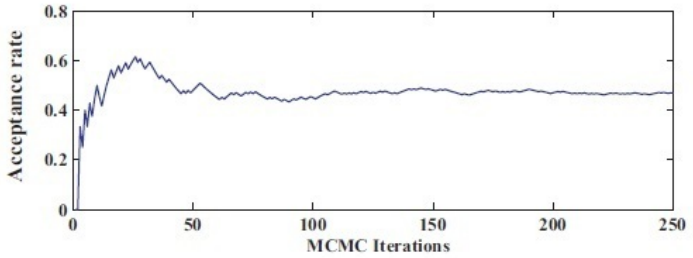
$$\psi_H\left(\rho^t,\rho^*\right) = \min\left[1,\frac{p\left(\rho^*|\boldsymbol{\beta},\sigma\right)}{p\left(\rho^t|\boldsymbol{\beta},\sigma\right)}\right],$$

der er sandsynligheden for, at det nye ρ^* accepteres i stedet for at beholde den nuværende værdi ρ^t. Det ønskes, at parameteren ρ genereres fra den tætte del af fordelingen og ikke sidder fast ude i enderne af fordelingen, hvor sandsaynlighedsmassen er lav, hvilket sikres ved at justere c i (6.19).

Eksempel 6.4:

Det ønskes ofte, at acceptområdet givet ved $\psi_H(\rho^t,\rho^*)$ ligger tæt på 50%, da traceplottet for data da hverken springer meget eller lidt, det vil sige, at der er bedre grundlag for grafisk at vurdere, om de simulerede værdier for ρ tilhører en ligevægt. Derfor kan

acceptområdet $\psi_H(\rho^t, \rho^*)$ vælges til at ligge mellem 40% og 60%. Hvis sandsynligheden i $\psi_H(\rho^t, \rho^*)$ bliver mindre end 40% for flere simulerede værdier i træk, ændres c til $c_1 = \frac{c}{1{,}1}$, hvilket medfører, at $\rho_1^* = \rho^{t_1} + c_1 \cdot N(0,1)$ vil være tættere på ρ^t, og dermed at acceptområdet $\psi_H(\rho^t, \rho^*)$ bliver større. På samme måde ændres c til $c_1 = 1{,}1 \cdot c$, hvis $\psi_H(\rho^t, \rho^*)$ for gentagne simuleringer af ρ er større end 60%, så sandsynligheden for accept bliver mindre. Det er som regel nødvendigt at justere c, da for eksempel små datasæt giver stor spredning i den betingede fordeling for ρ, mens det modsatte er tilfældet for store datasæt. Graden af justering er derfor ikke den samme for alle datasæt, selvom den anvendte metode er det, og i praksis prøver man sig ofte frem. Det ses af figur 6.1, at efter et burn-in på omkring 50 gennemløb af MCMC med M-H, hvor der også kan være ændret på parameteren c, befinder de resterende simuleringer sig i en ligevægt med $\psi_H(\rho^t, \rho^*)$ mellem 40% og 60%.



Figur 6.1: Acceptsandsynlighed ved MCMC, [LeSage and Pace, 2009, s. 138].

Hvis det vælges, at $\psi_H\left(\rho^t, \rho^*\right)$ skal være mellen 40% og 60%, vil der kun blive accepteret værdier for ρ halvdelen af tiden, og dermed er det nødvendigt at anvende MCMC flere gange for at have nok information til at danne en posteriorfordeling. □

6.4 MCMC-algoritmen

Det er muligt at udregne en $(q \times 1)$-vektor ud fra q tilfældigt udvalgte værdier af $\rho \in [0,1)$ til beregning af logdeterminantleddet $\ln\left|I_n - \rho W\right|$. Vektoren

$$\left[\ln\left|I_n - \rho_1 W\right| \quad \ln\left|I_n - \rho_2 W\right| \quad \ldots \quad \ln\left|I_n - \rho_q W\right|\right]^T$$

er en del af $p(\theta|\mathcal{D})$ i M-H-metoden, så regneregler til forenkling af udregning af determinanten hjælper også her. Det kan sikres, at $\rho \in [0,1)$ ved at afvise alle værdier af ρ uden for intervallet, når de genereres ved hjælp af M-H-sampling. Posteriorsandsynligheden for, at ρ er i det angivne interval, kan måles ved at notere, hvor mange simuleringer, der afvises.

For SAR-modellen fungerer MCMC ved at simulere fra tre fordelinger med arbitrære begyndelsesværdier givet for parametrene $\boldsymbol{\beta}$, σ^2 og ρ givet ved $\boldsymbol{\beta}_{(0)}$, $\sigma^2_{(0)}$ og $\rho_{(0)}$. Den simultane priorfordeling for $\boldsymbol{\beta}$ og σ^2 antages stadig at være $NIG(\mathbf{c}, \Sigma, a, b)$. De tre trin i simuleringen er som følger:

Trin 1
Vektoren $\boldsymbol{\beta}_{(1)}$ simuleres fra $p\left(\boldsymbol{\beta}|\sigma^2_{(0)},\rho_{(0)}\right)$ fra normalfordelingen

$$N\left(\mathbf{c}^*,\sigma_{(0)}\Sigma^*\right),$$

hvor

$$\Sigma^* = \left(X^T X + \Sigma^{-1}\right)^{-1}$$
$$\mathbf{c}^* = \Sigma^*\left(X^T\left(I_n - \rho_{(0)}W\right)\mathbf{y} + \Sigma^{-1}\mathbf{c}\right).$$

Den simulerede vektor $\boldsymbol{\beta}_{(1)}$ erstatter nu $\boldsymbol{\beta}_{(0)}$.

Trin 2
Parameteren $\sigma^2_{(1)}$ simuleres fra den inverse gammafordeling

$$p\left(\sigma^2|\boldsymbol{\beta}_{(1)},\rho\right) \sim IG\left(a^*,b^{**}\right),$$

hvor

$$a^* = a + \frac{n}{2}$$
$$b^{**} = b + \frac{\left(\left(I_n - \rho_{(0)}W\right)\mathbf{y} - X\boldsymbol{\beta}_{(1)}\right)^T\left(I_n - \rho_{(0)}W\right)\mathbf{y} - X\boldsymbol{\beta}_{(1)}}{2}.$$

Den simulerede værdi $\sigma^2_{(1)}$ erstatter herefter $\sigma^1_{(0)}$.

Trin 3
Simuler $\rho_{(1)}$ fra fordelingen $p\left(\rho|\boldsymbol{\beta}_{(1)},\sigma^2_{(1)}\right)$ ved brug af M-H-sampling.

De tre trin gennemgås mange gange, indtil et såkaldt burn-in, hvorefter mængden af de simulerede parametre anvendes til at lave posteriorestimater og inferens. Resten af de simulerede værdier for ρ efter burn-in er approksimeret konvergeret mod posteriorfordelingen. For at sikre sig at algoritmen har en ligevægtstilstand, kan man i praksis lave to gennemløb af algoritmen med forskellige startværdier. For eksempel kan laves først et kort gennemløb på 2500 gange, hvor de første 500 tages fra som burn-in, og derefter et langt gennemløb på for eksempel 8000 med de første 3000 estimater som burn-in. Hvis varians og middelværdi for de to gennemløb ligner hinanden, konvergerer MCMC-algoritmen, og de simulerede værdier efter burn-in vil da være estimaterne fra posteriorfordelingen. Herefter laves modeltest på de simulerede parametre for eksempel i form af hypotesetest og resiualer.

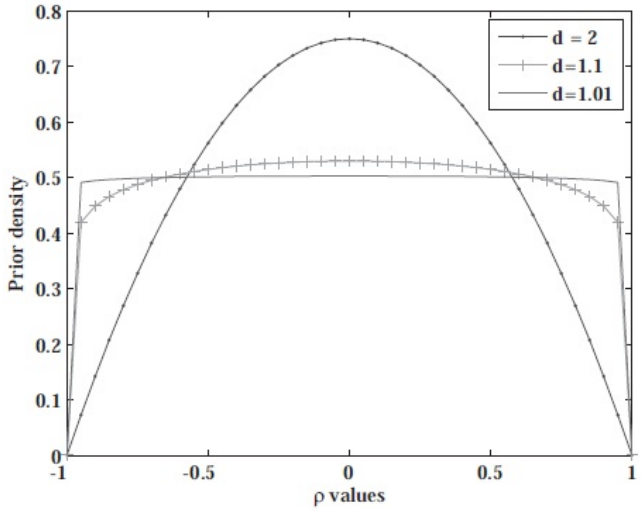
Det skal nævnes, at det viser sig, at simple, rumlige modeller som SAR-modellen generelt ikke har de store problemer med konvergens.

Eksempel 6.5:

Bayesiansk MCMC giver for store datasæt og bestemt valg af priorfordelinger samme estimater og inferens som ved maximum likelihood-estimation. Det er tilfældet, når ρ, $\boldsymbol{\beta}$ og σ^2 har priorfordelinger, der ikke indeholder ret meget information. Parameteren ρ kan for eksempel antages at have en priorfordeling givet ved

$$\pi\left(\rho\right) = \frac{1}{B(d,d)}\frac{(1+\rho)^{d-1}(1-\rho)^{d-1}}{2^{2d-1}} = \frac{1}{\int_0^1 t^{d-1}(1-t)^{d-1}dt}\frac{(1+\rho)^{d-1}(1-\rho)^{d-1}}{2^{2d-1}},$$

der er betafordelingen på intervallet $(-1,1)$, hvor $B(a,b) = B(d,d)$ er betafunktionen (se appendiks B.5), og konstanten $d > 0$. Priorfordelingen for ρ er en relativt uniform fordeling, der ikke indeholder væsentlig information, med en middelværdi på nul og er et alternativ til en uniform priorfordeling med $\rho \in (-1,1)$. Det ses på figur 6.2, at når d er tæt på nul, fås en priorfordeling, der indeholder relativt meget information, idet kurven på figur 6.2 da vil koncentrere sig mere om $\rho = 0$. Når $d = 1$, haves der en uniform fordeling.



Figur 6.2: Beta(d,d) som priorfordelingen for ρ [LeSage and Pace, 2009, s. 143].

Her er anvendt en multivariat normal priorfordeling for $\boldsymbol{\beta}$ med middelværdi nul og en diagonal varians-kovarians-struktur baseret på en ret stor variansskalar på $\sigma^2 = 10^{12}$, således at priorfordelingen ikke indeholder særlig information. Vi har her en priorfordeling, hvor variansen $\sigma^2 \to \infty$, det vil sige en uegentlig priorfordeling, hvilket måske kan give komplikationer, så det skal undersøges, hvorvidt der fås en veldefineret posteriorfordeling. Priorfordelingen for σ^2 er baseret på den inverse gammafordeling $IG(0,0)$, hvilket gør, at priorfordelingen ikke indeholder nogen særlig information. Der er hermed tale om to ikke-informative priorfordelinger. □

Ved sådanne priorvalg kan det vises, at estimaterne er de samme som ved maximum likelihood-estimation (se [LeSage and Pace, 2009, s. 142-145]).

6.5 Estimater for gennemsnitlige påvirkninger

MCMC kan anvendes til at danne posteriorfordelinger for funktioner af vigtige parametre, hvilket gør det simplere at undersøge sammenhænge mellem parametre.

Eksempel 6.6:
Et eksempel er en hypotese

$$\gamma = \alpha \cdot \beta < 1,$$

hvor α og β er parametre i en model, der er estimeret med MCMC. De m estimerede værdier for $\alpha_{(j)}$ og $\beta_{(j)}$ kan da producere $\gamma_{(j)}$ for $j = 1,\ldots,m$, der anvendes til at finde

posteriorsandsynligheden for at $\gamma < 1$. Sandsynligheden, for at $\gamma < 1$ er opfyldt, er givet ved andelen af estimaterne for γ, der er mindre end 1. Hvis der for eksempel genereres $m = 10.000$ estimater for γ, og 9.500 af dem er mindre end 1 er sandsynligheden for $\gamma < 1$ lig 95%. □

Der kan også laves en posteriorfordeling for hver af de gennemsntlige påvirkninger fra definition (2.16), (2.17) og (2.18). Hver gang, der laves et nyt estimat for parametrene ρ, σ^2 og $\boldsymbol{\beta}$ ved hjælp af MCMC, udregenes den tilhørende værdi for henholdsvis den gennemsnitlige totale, direkte og indirekte påvirkning ved at indsætte de estimerede parametre direkte i

$$\bar{M}(r)_{total} = \frac{\iota_n^T S_r(W)\iota_n}{n} = \left(1-\rho\right)^{-1}\beta_r \tag{6.20}$$

$$\bar{M}(r)_{direkte} = \frac{tr(S_r(W))}{n} = n^{-1}\sum_{j=0}^{q}\rho^j \operatorname{tr}(W^j)\boldsymbol{\beta} \quad \text{for} \quad q \to \infty \tag{6.21}$$

$$\bar{M}(r)_{indirekte} = \bar{M}(r)_{total} - \bar{M}(r)_{direkte}$$

for SAR-modellen, hvor (6.20) er fra sætning 2.19, mens (6.21) er resultatet fra (2.16). Det gør, at der til sidst haves en posteriorfordeling for hver af de gennemsnitlige påvirkninger. MCMC kan herved give mere præcise resultater end ved brug af maximum likelihood-estimation, når der haves en mindre mængde observationer.

6.6 Estimering af SAR-modellen med to vægtmatricer

Hvis SAR-modellen har to vægtmatricer W_1 og W_2 som i definition 3.7, fås loglikelihood-funktionen med fast $\boldsymbol{\beta}$ og σ^2 til

$$p\left(\mathbf{y}|\boldsymbol{\beta},\sigma^2,\rho\right) = -\frac{n}{2}\ln(2\pi\sigma^2) + \ln\left|I_n - \rho_1 W_1 - \rho_2 W_2\right| - \frac{e^T e}{2\sigma^2}$$

hvor $e = \mathbf{y} - (\rho_1 W_1 + \rho_2 W_2)\mathbf{y} - X\boldsymbol{\beta}$, da SAR-modellen med én vægtmatrix er normalfordelt. Når $\rho W = \rho_1 W_1 + \rho_2 W_2$ i definition 2.5 fås en normalfordeling. Den betingede fordeling for parametrene ρ_i, $i = 1,2$, har formen

$$p\left(\rho_i|\boldsymbol{\beta},\sigma\right) \propto \frac{p\left(\rho_i,\beta,\sigma|\mathbf{y}\right)}{p\left(\beta,\sigma|\mathbf{y}\right)}.$$

Likelihoodfunktionerne er givet ved det samme som før blot med en anden determinant givet ved $|I_n - \rho_1 W_1 - \rho_2 W_2|$ jævnfør definition 3.7, og en løsning af ovenstående vil kræve multivariate løsningsmetoder samt estimering af både ρ_1 og ρ_2.

6.7 Central posterior interval

Dette afsnit er skrevet ud fra [Lee, 2004, afsnit 2.6].

I forbindelse med bayesiansk statistik anvendes ofte begreberne Highest density region (HDR) og central posterior interval (CPI). Når man har gennemført et eksperiment, så

er al relevant informationen indeholdt i posterioren. Det kan ofte være af interesse, at fortolke posterioren ved at angive et område eller interval hvori den ukendte variabel befinder sig med en given sandsynlighed. For et givet $\alpha \in (0,1)$ kan man således angive det mindste område, så arealet under tæthedsfunktionen er lig med α. Et sådan område kaldes en highest density region (HDR) og er udtrykt i definition 6.7.

Definition 6.7 (Highest density region):
Et α HDR er området Θ, hvor

$$\int_{\Theta} p(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta} = \alpha, \qquad \alpha \in (0,1),$$

og $\sup_{\tilde{\boldsymbol{\theta}} \notin \Theta} p(\tilde{\boldsymbol{\theta}}|\mathbf{Y}) \leq p(\boldsymbol{\theta}|\mathbf{Y})$ for alle $\boldsymbol{\theta} \in \Theta$. Værdien α angiver signifikansniveauet.

En ulempe ved HDR er, at HDR ikke nødvendigvis er et sammenhængende interval. I stedet for HDR kan anvendes CPI, der er invariant under transformation og altid giver et sammenhængende interval. Et CPI er imidlertid kun defineret for endimensionale variable, hvorimod HDR også kan anvendes for flerdimensionale variable.

Definition 6.8 (Central posterior interval):
Et α CPI er givet ved området mellem $\frac{1-\alpha}{2}$-fraktilen og $1-\frac{1-\alpha}{2}$-fraktilen, hvor $\alpha \in (0,1)$.

Størrelserne HDR og CPI er lig hinanden, hvis posteriorfordelingen er symmetrisk og har et globalt maksimum som eneste kritiske punkt, og derfor kan vi bruge begge begreber, hvis posteriorfordelingen for eksempel er normalfordelt. Størrelserne HDR og CPI kan også være ens i nogle specialtilfælde, hvor der er flere kritiske punkter. Der ses nu på et eksempel på, hvordan CPI kan udregnes.

Eksempel 6.9:
Det ønskes at bestemme et 95%-CPI for en standard normalfordeling, så der er 95% sandsynlighed for, at θ er i intervallet $[-x,x]$. Hermed skal de to værdier x og $-x$ bestemmes, så fordelingsfunktionen for standard normalfordelingen, Φ, giver $\Phi(x) - \Phi(-x) = 0.95$. Det haves, at

$$\Phi(x) - \Phi(-x) = \Phi(x) - (1 - \Phi(x)) = 2\Phi(x) - 1 = 0.95,$$

så

$$x = \Phi^{-1}\left(\frac{1.95}{2}\right) = \Phi^{-1}(0.975) = 1.96,$$

hvor sidste lighed bestemmes ved hjælp af en tabel (se [Olofsson, 2005, p. 468]) for fordelingsfunktionen for standard normalfordelingen. Dermed er 95% af sandsynlighedsmassen i intervallet $[-1{,}96; 1{,}96]$. □

Databehandling 7

Vi anvender et datasæt med alle bolighandler fra HOME-kæden i Danmark i perioden fra januar 2002 til og med januar 2009, hvor der er noteret 96.010 observationer i alt fra 96 kommuner. Vi vil gerne undersøge sammenhængen mellem boligers beliggenhed i forskellige kommuner med deres boligpris, og hvordan en sammenhæng mellem priserne kan beskrives med en SAR-model. Vi vil se på den klassiske metode og sammeligne den med den bayesianske metode ved at tilpasse en model fra begge, og der ses blandt andet på parameterestimater, p-værdier, konfidensintervaller og CPI'er for begge metoder.

7.1 Oplysninger om datasæt

Det er nødvendigt at foretage en udrensning og bearbejdning af registreret data på grund af tastefejl, irrelevante baggrundsvariable, manglende dataoplysninger til en del relevante observationer, mellemrum mellem tal, ikke konsistent og entydig notation af blandt andet afstandsangivelser, hvor der er anvendt forskellige mål ved, at nogle data for eksempel er noteret i meter, mens andre er noteret i km. Vi har bearbejdet data, så alle afstande angives i meter, redigeret tastefejl så vidt muligt og fjernet alle irrelevante baggrundsvariable. I datasættet er der observationer fra i alt 96 kommuner, men dette reduceres til 94 forskellige kommuner i alt, efter at vi har slettet alle observationer med manglende oplysninger, hvilket vil sige, at rækker med boliger med NA-værdier i baggrundsvariable. Desuden har vi udelukket kommunen Albertslund, idet der efter udrensning af data kun var en enkelt observation tilbage fra denne kommune. Antallet af observationer i alt i datasættet reduceres efter udrensning til 25.586 antal observationer i alt. For at kunne lave naborelationer mellem kommunerne er der desuden taget gennemsnit af værdierne for hver kommune, så antallet af observationer i alt er 94.

Datasættet indeholder 44 baggrundsvariable i alt, men vi får langt fra brug for dem alle, da ikke alle baggrundsvariable er relevante for netop vores undersøgelse. Da vi hovedsageligt vil fokusere på boligpriser set i forhold til beliggenhed, er en del baggrundsvariable fra det oprindelige datasæt udeladt såsom tagdækning, ydervægge, nye vinduer, nyt badeværelse, garage/carport, salgsdato, vurderingsår osv. De baggrundsvariable, vi har valgt at medtage i vores model, er følgende: Kommune, grundareal, boligareal, afstand til offentlig transport, boligtilstand og kvadratmeterpris, der er udregnet ved hjælp af kontantpriserne. Idet boligens stand og areal har meget at sige for boligens pris, har vi specifikt valgt at medtage disse baggrundsvariable i vores model. Vi ville også gerne have inkluderet afstand til strand i vores model, men da datasættet er meget mangelfuldt med hensyn til oplysning af det, så har vi valgt at udelade dette som baggrundsvariabel.

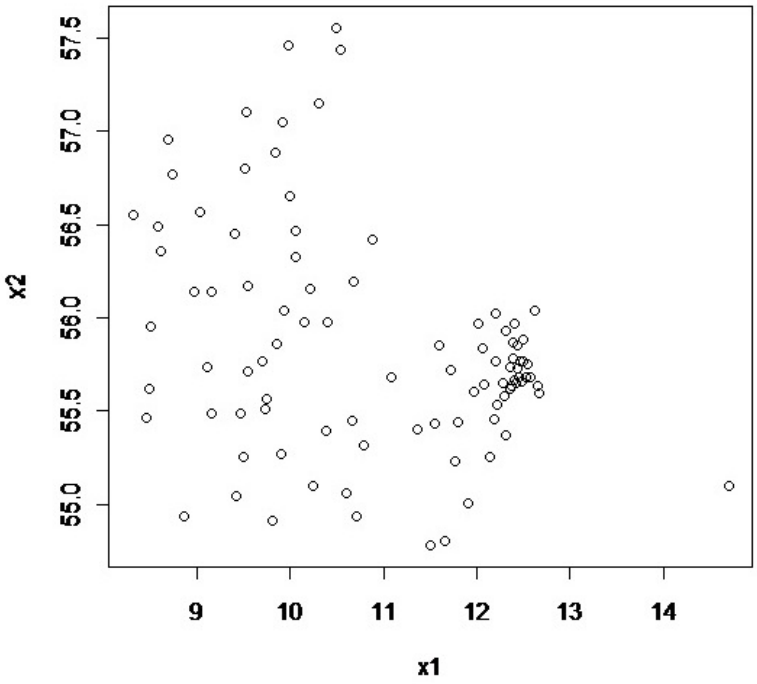
Boligens stand er angivet ud fra de tre vurderinger god=1, middel=2 og dårlig=3. HOME inddeler ejendomstyper i to store grupper, villa og ejerlejlighed, kaldet ejendomskatego-

rier. I projektet har vi valgt at se bort fra inddelingen af ejendomstyper og i stedet medtaget grundareal, der indikerer, om der er tale om en lejlighed eller et hus, da lejligheder har et grundareal på nul.

7.2 Konstruktion af Voronoi-diagram over observationer i datasæt

Vi har valgt at dele Danmark op i kommuner og sammenligne de gennemsnitlige boligpriser mellem kommunerne. Naborelationer udtrykt ved matricen W i SAR-modellen laves ud fra et Voronoi-diagram, som er omtalt i afsnit 2.4, over kommunerne ud fra længde- og breddegrader for disse. Længde- og breddegrader for alle kommuner bliver hermed afgørende for hver kommunes position i forhold til hinanden. Længde- og breddegrader er fundet ved hjælp af internetsiden [Wick], og vi har anvendt enten den største eller mest centrale by i hver kommune til at bestemme én længdegrad og én breddegrad for hver kommune. Et kort over alle kommuner ses på figur 7.1, og på figur 7.2 ses en graf, hvor kommunerne i datasættet er angivet ved hjælp af længde- og breddegrader.

Figur 7.1: Kort over alle kommuner i Danmark [Jensen].



Figur 7.2: Kommunerne angivet ved hjælp af længde- og breddegrader.

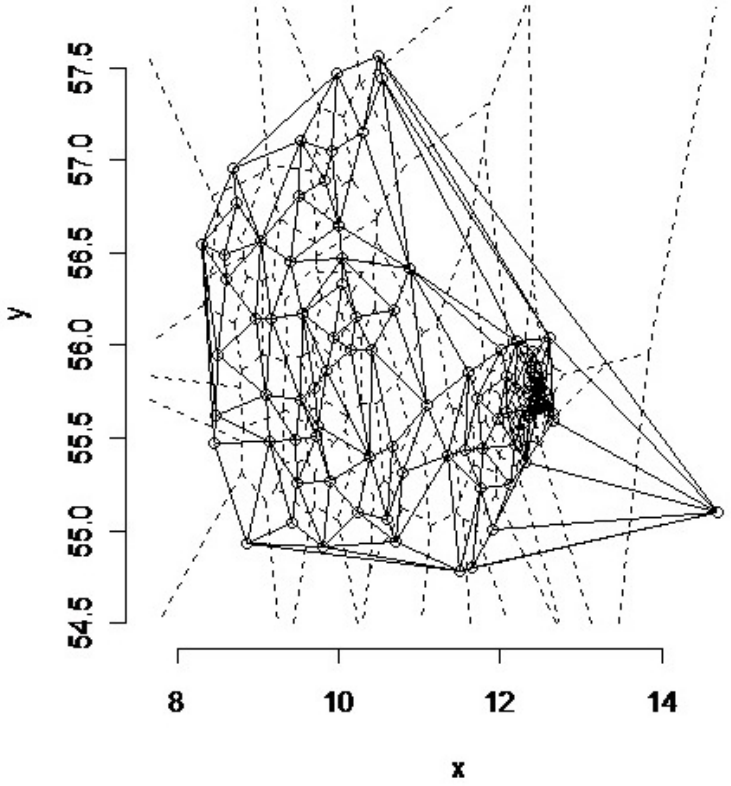
Kode 7.1 laver et Voronoi-diagram over længde- og breddegrader for de 94 kommuner fra datasættet og definerer derudover nabomatricen W.

```
1  data<-read.table("voronoi.csv", sep=";", header=T) #indlæser længde- og
       breddegrader for alle kommuner
2
3  x1 <- data[,2] # definerer x som vektor med breddegrader
4  x2 <- data[,1] # definerer y som vektor med længdegrader
5
6  plot(x1,x2) # sådan ser (x1,x2) ud som koordinater på et landområde
7
8  install.packages("deldir") #installerer pakke til Voronoi-diagrammer
9  library(deldir)
10
11 V = deldir(x1,x2) # Her laves et Voronoi-diagram
12
13 plot(V) # stiplede linjer angiver Voronoi-diagrammet, fuldtoptrukne linjer angiver nabopar og de
       stiplede linjer angiver Delaunay-trekanterne
14
15 V$delsgs # Viser information om nabostruktur
16
17 nb = V$delsgs[,5:6] # Udtager informationen af V om indeksnumrene på nabopar i 5. og 6.
       kolonne
18
19 W = matrix(0,94,94) # nabomatrix W med nuller
20 for (i in 1:dim(nb)[1]) W[nb[i,1],nb[i,2]] = 1 # 1-taller på nabopars indgange i en
       nedre triangulær matrix
21
22 W <- edit(W) # Manuel ændring af enkelte naboobservationer
23
```

```
24 W = W + t(W)  # Inkludering af den øvre triangulære matrix
25 W = W/apply(W,1,sum)  # W gøres rækkestokastisk
26
27 apply(W,1,sum)  # Her undersøges om rækkesummerne er 1
```

R-kode 7.1: Kode til at konstruere nabomatricen W.

Først indlæses længde- og breddegrader for alle kommuner og defineres som vektorer. Der laves et Voronoi-diagram, og indgangene i nabomatricen W tildeles et ettal, hvis de kommuner, som indgangene repræsenterer, er naboer. Herefter gøres W rækkestokastisk. Koden plotter også selve Voronoi-diagrammet, der er vist på figur 7.3.



Figur 7.3: Voronoi-diagram over kommuner.

Det kan ses, at koden har defineret, at Bornholm er nabo med blandt andet Læsø, hvilket ikke passer med virkeligheden, så der er foretaget en manuel redigering af matricen W ved naborelationer, som Voronoi-diagrammet har angivet som naboer, men i virkeligheden er adskilt af hav.

Dog er der en ulempe ved anvendelse af Voronoi-diagrammet i tilfældet med danske kommuner. Afgrænsningerne for kommunerne i Danmark har nemlig meget forskellige og atypiske former, i forhold til hvis man for eksempel ser på afgrænsningerne af stater i USA. Man kan derfor forestille sig, at der vil være nogle kommuner, der ikke vil blive noteret som naboer, selvom de i virkeligheden er det. For eksempel kan det ske at Ikast-Brande kommune (se stjernen på figur 7.1) kun noteres som værende nabo med kommunerne Silkeborg, Herning, Horsens og Vejle, hvorimod

naboinformationen om Billund, Hedensted og Viborg måske vil gå tabt. Man får mere præcise naborelationer, hvis naborelationerne noteres manuelt i hånden, så man er sikker på, at alle naborelationer noteres korrekt.

7.3 Konstruktion af SAR-model for data

Vi vil nu se på, hvilke elementer i datasættet, der udgør de forskellige led i SAR-modellen givet ved datavektoren $\mathbf{y}$, designmatricen X og nabomatricen W.

Datavektoren $\mathbf{y}$ udtrykker kontantprisen for en bolig divideret med boligarealet for boligen, således at der fås en kvadratmeterpris på alle boliger for hver kommune. Som tidligere beskrevet skelner datasættet mellem boligernes type, men vi har lavet en forsimplet udgave, idet vi har valgt ikke at skelne mellem lejligheder og huse, og dermed har vi udregnet en samlet, gennemsnitlig kvadratmeterpris for alle boligtyper for hver kommune. Det er et meget forsimplet billede af virkeligheden, men der kompenseres en smule for dette, idet grundareal for hver bolig inddrages i baggrundsvariablene gennem designmatricen X. Herigennem tages der indirekte højde for lejlighedstypen, idet grundarealet for for eksempel en lejlighed må forventes at være tæt på nul og større for huse. Vi tager hermed gennemsnittet af alle kvadratmeterpriserne for alle boliger i hver kommune og lader disse værdier repræsentere hver kommune, således at der er én gennemsnitlig kvadratmeterpris per kommune.

Designmatricen X i SAR-modellen indeholder alle relevante baggrundsvariable givet ved grundareal, afstand til centrum, afstand til offentlig transport, boligtilstand og boligareal.

Nabomatricen W er naturligvis udtrykt ved Voronoi-diagrammet fra afsnit 7.2.

7.4 Tilpasning af SAR-modellen

Nu tilpasses en SAR-model til datasættet fra HOME. Vi bruger pakken *spdep* i R og anvender kommandoen *lagsarlm*, der anvender maximum likelihood-estimation til estimering af rumlige, simultane, autoregressive, laggede modeller på formen $\mathbf{y} = \rho W\mathbf{y} + X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Parameteren ρ findes først ved brug af kommandoen *optimize()*, hvorefter $\boldsymbol{\beta}$ og andre modelparametre findes ved mindste kvadraters metode, hvilket svarer til maximum likelihood-estimation, idet ρ er kendt. Den samlede kode kan ses i kode 7.2.

```
1 install.packages("spdep") # installerer pakke til SAR-model
2 library(spdep)
3
4 data1<-read.table("homegennemsnit.csv", sep=";", header=T) #datasæt med
      relevante baggrundsvariable
5 Wl = mat2listw(W, style="W") # W konverteres til brugbar form i SAR-modellen
6 eigenw(mat2listw(W, style="W")) # Her vises egenværdierne
7
8 SAR <- lagsarlm(Kvadratmeterpris~Afstand.til.centrum + Boligareal +
      Afstand.off.transport + Boligtilstand, data=data1, method="MC",
      listw = Wl) # Her tilpasses SAR-modellen til datasættet
9 # method= "MC" betyder, at der anvendes MC-approksimation til estimering af log-determinanten
```

R-kode 7.2: Kode til at tilpasse en SAR-model til data.

Kode 7.2 giver blandt andet egenværdierne for SAR-modellen, der er vist på figur 7.4, hvor det ses, at den maksimale egenværdi er 1, og at alle egenværdierne er reelle tal.

```
 [1]  1.000000000  0.981364509  0.931017703  0.885097556  0.875941336
 [6]  0.822949320  0.793585684  0.767448942  0.758592776  0.698325674
[11]  0.666409108  0.642602454  0.594911701 -0.559228995  0.558691635
[16] -0.550149997 -0.523772250  0.522430976  0.495925646 -0.487254501
[21] -0.478735965 -0.467484948 -0.457145600  0.456808499 -0.449642381
[26] -0.445363579 -0.443103734 -0.439642056  0.437265884 -0.430148832
[31] -0.424447990 -0.418360963  0.417301311 -0.412914439 -0.406952348
[36] -0.402635533 -0.396278969 -0.385472866 -0.381380282  0.373865218
[41] -0.372493065  0.372352153 -0.371405512  0.361557015 -0.359854793
[46] -0.353793193 -0.346087822 -0.344350267 -0.340893692  0.338534375
[51] -0.332733382 -0.325411778 -0.317471920 -0.302804395  0.301397987
[56] -0.294864503 -0.288770689 -0.280179598  0.276535869  0.263581974
[61] -0.263543325 -0.250593144 -0.242793998 -0.235723950 -0.233430760
[66] -0.222705317 -0.215905385  0.212729334 -0.197761770  0.196269716
[71] -0.192068023 -0.185540595  0.179972424 -0.168937803  0.163252311
[76] -0.152144661 -0.143893648  0.128833944 -0.125519825  0.115416130
[81] -0.102663400  0.097186411 -0.094891269 -0.087903519  0.082200714
[86] -0.073003892 -0.066220667  0.065204529  0.053054259 -0.040774908
[91]  0.022840201 -0.021756399 -0.007252664  0.004804481
```

Figur 7.4: Egenværdier for W i SAR-modellen.

Idet der ikke er tale om et særlig stort datasæt, fordi der kun er 94 kommuner i det tilpassede datasæt, så er det i dette tilfælde ikke strengt nødvendigt at estimere sporet, som det er beskrevet i kapitel 4.

I kode 7.2 ses det, at der er medtaget færre baggrundsvariable end tidligere nævnt. Det skyldes, at mange baggrundsvariable viser sig ikke at være signifikante, hvis alle de tidligere udvalgte tages med. Ved at eksperimentere med at fjerne nogle af variablene, viser det sig, at de variable, der er med i kode 7.2, giver en p-værdi omkring 5%. Hermed udgår for eksempel baggrundsvariablen grundareal, idet den i modstrid med de indledende overvejelser ikke er signifikant for vores datasæt.

Vi ønsker at trække en række oplysninger om SAR-modellen ud af den tilpassede model i kode 7.2, hvilket gøres i kode 7.3. Kode 7.3 giver estimater for $\boldsymbol{\beta}$, σ^2 og ρ og derudover to testværdier i form af likelihood ratio-testen og Wald-testen.

```
summary(SAR)  # Her vises en del information om SAR-modellen anvendt på datasættet (blandt
    andet en z-test og en Waldtest)

names(summary(SAR))  # Her ses al den information, der kan trækkes ud af modellen
coef(summary(SAR))  # Her ses estimater for både rho og beta
impacts(SAR, listw = Wl)  # Her ses de gennemsnitlige påvirkninger

residuals(SAR)  # Residualer findes og plottes
plot(residuals(SAR))
```

R-kode 7.3: Kode til at finde information om SAR-modellen.

Outputtet for *summary(SAR)* i kode 7.3 ses på figur 7.5.

```
Call:lagsarlm(formula = Kvadratmeterpris ~ Afstand.til.centrum + Boligareal +
    Afstand.off.transport + Boligtilstand, data = data1, listw = Wl,
    method = "MC")

Residuals:
     Min       1Q   Median       3Q      Max
-6981.49 -1932.45  -198.46  1910.76  6858.82

Type: lag
Coefficients: (numerical Hessian approximate standard errors)
                        Estimate  Std. Error z value  Pr(>|z|)
(Intercept)           18131.13286  2510.47388  7.2222 5.116e-13
Afstand.til.centrum       0.29857     0.10346  2.8859  0.003903
Boligareal              -98.42586    15.65279 -6.2881 3.214e-10
Afstand.off.transport    -2.96672     1.93300 -1.5348  0.124838
Boligtilstand         -1044.62470   594.53253 -1.7571  0.078909

Rho: 0.68769, LR test value: 74.816, p-value: < 2.22e-16
Approximate (numerical Hessian) standard error: 0.058054
    z-value: 11.846, p-value: < 2.22e-16
Wald statistic: 140.32, p-value: < 2.22e-16

Log likelihood: -884.545 for lag model
ML residual variance (sigma squared): 7701300, (sigma: 2775.1)
Number of observations: 94
Number of parameters estimated: 7
AIC: 1783.1, (AIC for lm: 1855.9)
```

Figur 7.5: Output for SAR-modellen.

Det ses heraf, at estimaterne for modelparametrene $\boldsymbol{\beta}$, σ^2 og ρ er givet ved

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 18.131{,}13 \\ 0{,}2986 \\ -98{,}43 \\ -2{,}9667 \\ -1044{,}62 \end{bmatrix}$$

$$\hat{\sigma}^2 = 7701300$$

$$\hat{\rho} = 0{,}6877.$$

Den første indgang i $\hat{\boldsymbol{\beta}}$ givet ved $\hat{\beta}_1$ er skæringen. Den har typisk ikke nogen praktisk fortolkning, da den er kvadratmeterprisen, når alle baggrundsvariable er sat lig nul. Den anvendes dog grafisk til at beskrive kurven for $\mathbf{y}$, og der er ikke nogen logisk begrundelse for, at grafen nødvendigvis skal starte i nul, hvilket den vil gøre, hvis skæringen er nul.

Desuden giver det intutivt ikke mening at have en bolig på nul kvadratmeter, hvilket vil sige en bolig, der ikke eksisterer.

Afstanden til centrum i $\hat{\beta}_2$ påvirker boligprisen positivt. Det kan virke lidt ulogisk, da man måske vil forvente, at det er dyrest at bo midt i centrum, da man her er tæt på faciliteter såsom skoler, butikker og indkøbscentre mm. Der kan dog være flere årsager til, at det kan være dyrere at bo en vis afstand fra centrum, for eksempel kan der være ønsker om have og naturområder, som opfyldes bedre uden for centrum, og forstaden er måske mere børnevenlig med hensyn til trafik og støj. Derudover er der en middelløsning mellem at bo lige midt i centrum og at bo på for eksempel et husmandssted langt uden for større byområder, hvilket kunne være en forstad, hvor det tit er dyrere at bo, idet man har større huse med have og alligevel er tæt på byen. Da afstand til centrum kun angiver en samlet værdi, kan det være svært at få det nyancerede billede af denne problemstilling med i modellen.

Boligarealet i $\hat{\beta}_3$ påvirker kvadratmeterprisen negativt, således at kvadratmeterprisen bliver mindre des større boligen er. Dette virker logisk, idet store boliger ellers vil være meget dyre, og det må antages, at man på et tidspunkt rammer en øvre grænse. For eksempel er forskellen på, om man har en bolig på 275 eller 300 kvadratmeter ikke så stor, når man er oppe i den størrelsesorden af boliger, men hvis kvadratmeterprisen er 8.000 kr. som for visse boliger i datasættet, så er forskellen på boligprisen 200.000 kr. for kun 25 kvadratmeter.

Afstand til offentlig transport i $\hat{\beta}_4$ påvirker boligprisen negativt, således at des længere der er til offentlig transport des lavere er kvadratmeterprisen for en bolig, hvilket udtrykker et behov blandt boligejere for hurtig adgang til offentlig transport. Det afspejler et generelt behov for offentlig transport, uanset om man bor i eller uden for centrum. Selvom behovet for offentlig transport må formodes at være mindre for dem, der vælger at bosætte sig uden for byområder, idet de ofte har bil, så ønsker de fleste stadigvæk at have offentlig transport i nærheden, så for eksempel børn kan komme i skole, eller der haves et alternativ til bilen.

Boligtilstanden $\hat{\beta}_5$ påvirker boligpriserne negativt, hvilket hænger sammen med, at boligerne er inddelt i kategorierne 1, 2 og 3, hvor 1 er bedst og 3 er dårligst, så når boligtilstanden stiger i værdi, mindskes kvadratmeterprisen.

Det ses også, at $\hat{\rho}$ er positiv og mindre end 1, hvilket betyder, at der er positiv korrelation mellem huspriser i naboliggende kommuner.

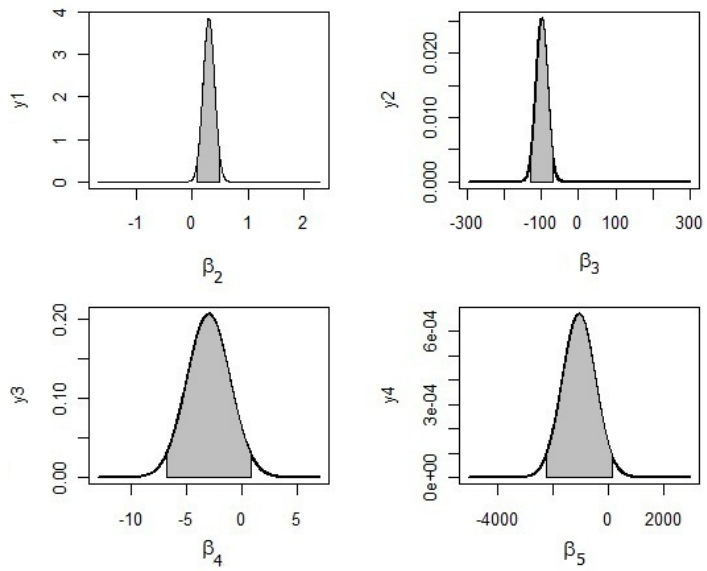
Parameteren $\hat{\sigma}^2$ virker høj, men hvis der ses på spredningen $\hat{\sigma} = 2.775{,}1$, er det en passende spredning for kvadratmeterprisen, da de gennemsnitlige kvadratmeterpriser i datasættet ligger mellem 2.818 og 29.999 kr..

Den approksimerede værdi for 97,5-fraktilen for normalfordelingen er givet ved 1,96, således at det jævnfør den centrale grænseværdisætning antages, at 95% af arealet under fordelingskurven for en parameter ligger inden for 1,96 standardafvigelser fra middelværdien. Vi kan derfor udregne et konfidensinterval for alle parametrene ved at addere maximum likelihood-estimatet (middelværdien) for hver parame-

ter med standardafvigelsen for denne multipliceret med 1,96. Det skal her bemærkes, at standardafvigelsen på figur 7.5 tager højde for kvadratroden af $n = 94$. Oplysningerne til udregning af konfidensintervallerne fås fra figur 7.5, så eksempelvis $\hat{\boldsymbol{\beta}}_2 \in [0{,}29857 + (1{,}96 \cdot 0{,}0346); 0{,}29857 + (1{,}96 \cdot 0{,}0346)] = [0{,}0958; 0{,}5014]$. Konfidensintervallerne for alle parametrene $\boldsymbol{\beta}$ og ρ er givet ved

$$\begin{aligned}
\beta_1 &\in [13.210{,}60; 23.051{,}66] \text{ med 95\% sikkerhed} \\
\beta_2 &\in [0{,}0958; 0{,}5014] \text{ med 95\% sikkerhed} \\
\beta_3 &\in [-129{,}10; -67{,}75] \text{ med 95\% sikkerhed} \\
\beta_4 &\in [-6{,}7554; 0{,}8220] \text{ med 95\% sikkerhed} \\
\beta_5 &\in [-2.209{,}91; 120{,}66] \text{ med 95\% sikkerhed} \\
\rho &\in [0{,}5739; 0{,}8015] \text{ med 95\% sikkerhed .}
\end{aligned}$$

Konfidensintervallerne er relativt små i forhold til de parametre, de indeholder, og det ses på figur 7.6, at størstedelen af sandsynlighedsmassen er koncentreret om de respektive middelværdier.



Figur 7.6: Fordelinger og konfidensintervaller for indgangene i $\boldsymbol{\beta}$.

Det undersøges, hvorvidt der er nogle af parametrene $\boldsymbol{\beta}$, σ^2 og ρ, der kunne være lig med nul, og derfor bør udgå af modellen, hvilket vil sige, at vi ser på hypotesen, hvor

$$\begin{cases} H_0: & \theta_i = 0 \\ H_1: & \theta_i \neq 0, \end{cases}$$

hvor der ses på parametrene $\theta_i = \beta_i$ for $i = 1, \ldots, 5$, $\theta_i = \sigma^2$ og $\theta_i = \rho$. Vi antager, at parametre med en p-værdi på 5% eller mindre er signifikante, hvilket vil sige, at de har betydelig indflydelse i modellen.

Det ses på figur 7.5, at der ikke er særlig stor sandsynlighed for, at nogle af parametrene i modellen skulle være lig med nul, idet p-værdien er meget lav for alle parametre undtaget

afstand til offentlig transport og boligtilstand. p-værdierne for afstand til offentlig transport og boligtilstand er henholdsvis 0,1248 og 0,078909, hvilket er større end 0,05. p-værdien for offentlig transport er væsentlig større end 5%, og man kan derfor overveje, om parameteren bør udgå fra modellen, hvilket understøttes af konfidensintervallet for parameteren, der indeholder nul. Ligeledes kan en udeladelse af boligtilstand overvejes på trods af, at konfidensintervallet ikke indeholder nul. En udeladelse af begge parametre vil dog resultere i en noget reduceret model. På figur 7.5 ses også, at boligarealet har meget høj signifikans, hvilket er naturligt, da boligpriserne netop er angivet i kvadratmeterprisen. Det ses desuden, at der er stor sandsynlighed for, at der er tale om rumlig afhængighed mellem observationerne, idet p-værdien for ρ er mindre end $2{,}22 \cdot 10^{-16}$ for både likelihood ratio-testen og Wald-testen, hvor det vides fra afsnit 5.3, at en LR-værdi tæt på nul forkaster nulhypotesen $\rho = 0$.

På figur 7.5 er også z-værdien for en hypotesetest af nulhypotesen $\rho = 0$. Jævnfør den centrale grænseværdisætning, så kan ρ antages at være normalfordelt, og de kritiske værdier for normalfordelingen ved 0.025 og 0.975-fraktilen er henholdsvis $-1{,}959964$ og 1,959964. Hermed er z-værdien $z = 11{,}846$ større end den kritiske værdi 1,959964, der er 0,975-fraktilen for normalfordelingen. Nulhypotesen forkastes dermed, og der er dermed stor sandsynlighed for, at $\rho \neq 0$. Samme konklusion kan drages ud fra Waldtesten, hvor der ses på fraktilerne for χ^2-fordelingen i stedet for normalfordelingen. På figur 7.5 ses det, at p-værdien er mindre end $2{,}22 \cdot 10^{-16}$, hvilket er mindre end det valgte konfidensniveau på 5%, så nulhypotesen forkastes, og vi kan antage, at $\rho \neq 0$.

De gennemsnitlige påvirkninger udregnes i kode 7.3 og ses på figur 7.7.

```
Impact measures (lag, exact):
                            Direct      Indirect         Total
Afstand.til.centrum      0.3424971     0.6135284     0.9560254
Boligareal            -112.9061581  -202.2532049  -315.1593630
Afstand.off.transport    -3.4031819    -6.0962526    -9.4994345
Boligtilstand         -1198.3087253 -2146.5771603 -3344.8858856
```

Figur 7.7: Gennemsnitlige påvirkninger.

Påvirkningerne kan ses som en form for hældning for de forskellige parametre. Vi kan jævnfør afsnit 2.5 se, at

$$\begin{aligned}
\mathbf{y} &= \sum_{r=1}^{k} (S_r(W)\mathbf{x}_r) + V(W)\boldsymbol{\varepsilon} \\
&= \sum_{r=1}^{k} \left(\left(I - \rho W\right)^{-1} \beta_r \mathbf{x}_r \right) + V(W)\boldsymbol{\varepsilon} \\
&= \sum_{r=1}^{k} \left(\left(I + \rho W + \rho^2 W^2 + \ldots\right) \beta_r \mathbf{x}_r \right) + V(W)\boldsymbol{\varepsilon},
\end{aligned}$$

hvor $V(W)$ er givet i (2.14), og $\boldsymbol{\varepsilon}$ er fejlvektoren. Derudover er β_r den r'te indgang i vektoren $\boldsymbol{\beta}$, og $\mathbf{x}_r$ er en søjle i designmatricen X. I denne model er de gennemsnitlige påvirkninger udtrykt gennem størrelsen $S_r(W) = \left(I + \rho W + \rho^2 W^2 + \ldots\right)$, så det haves, at

$I\beta_r\mathbf{x}_r$	er en vektor, der angiver direkte påvirkninger
$\rho W\beta_r\mathbf{x}_r$	er en vektor, der angiver påvirkninger på naboer
$\rho^2W^2\beta_r\mathbf{x}_r$	er en vektor, der angiver påvirkninger på naboers naboer
$\left(I+\rho W+\rho^2W^2+\ldots\right)\beta_r\mathbf{x}_r$	angiver alle påvirkninger i alt.

Sidstnævnte udtryk er ikke særlig anvendeligt i praksis, men hvis der divideres med antallet af observationer, fås at de gennemsnitlige, samlede påvirkninger er givet ved

$$n^{-1}\iota\left(I+\rho W+\rho^2W^2+\ldots\right)\beta_r\iota^T\mathbf{x}_r, \tag{7.1}$$

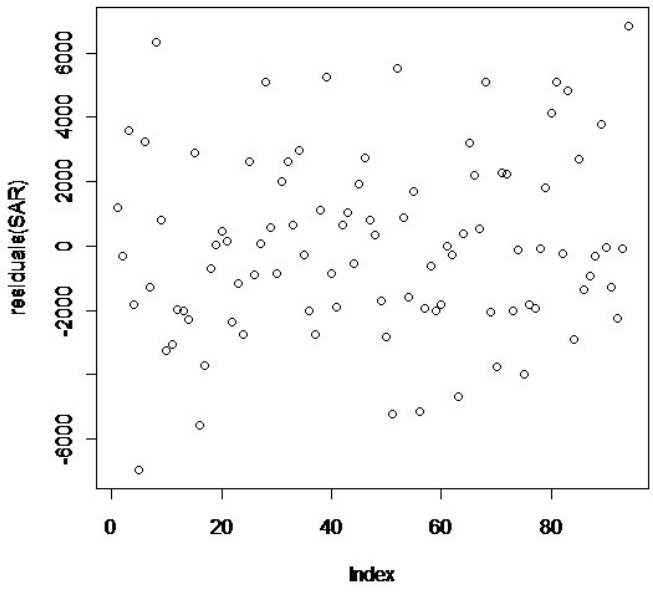
der angiver alle gennemsnitlige, samlede påvirkninger på sig selv og naboer af alle ordener. Samtidig er de gennemsnitlige, direkte påvirkninger givet ved

$$\mathrm{tr}\left(\left(I+\rho W+\rho^2W^2+\ldots\right)\beta_r\right)\mathbf{x}_r, \tag{7.2}$$

der angiver gennemsnitlige, direkte påvirkninger. Vi kan dermed se, at hvis vektoren $\mathbf{x}_r$ ændres til $\mathbf{x}_r+\delta$, hvilket på figur 7.7 kan svare til, at det gennemsnitlige boligareal øges på huse i alle kommuner, så haves, at ændringen i hvert boligareal påvirker kvadratmeterprisen for den kommune, der foretager en ændring i arealet. Det vil sige, der er en direkte påvirkning fra, at en kommune øger boligarealet, til den samme kommune oplever et fald i kvadratmeterprisen, og gennemsnittet af denne slags påvirkninger for hver kommune er et fald i kvadratmeterprisen på $\delta\cdot 202$ kr..

De gennemsnitlige, indirekte påvirkninger fås ved substraktion af de to udtryk i (7.1) og (7.2) og svarer til, at hvis for eksempel boligerne i en kommune hver øges med 1 kvadratmeter, så falder alle kvadratmeterpriser med cirka 113 kr. Hermed er de samlede, gennemsnitlige påvirkninger cirka lig med −315 kr., hvis husene i kommunerne øges med 1 kvadratmeter.

Residualerne udregnes i kode 7.3 og ses på figur 7.8.



Figur 7.8: Residualplot.

Generelt ser residualplottet ud til at have en jævn spredning omkring nul, og ses der på figur 7.5, ligger residualerne i intervallet $[-6981{,}49; 6858{,}82]$, og middelværdien er $-198{,}46$, hvilket ikke er en stor afvigelse fra nul, når man tager de generelle afvigelser i residualerne i betragtning. Ingen af residualerne ligger mere end tre standardafvigelser væk fra middelværdien, da $3 \cdot \sigma = 8325{,}3$, så der er ingen outliers i residualplottet. Det kan derfor konkluderes, at residualplottet tilsyneladende ikke afviser den tilpassede SAR-model for datasættet fra HOME, og at der ikke er udeladt betydningsfulde baggrundsvariable.

7.5 Tilpasning af den bayesianske SAR-model

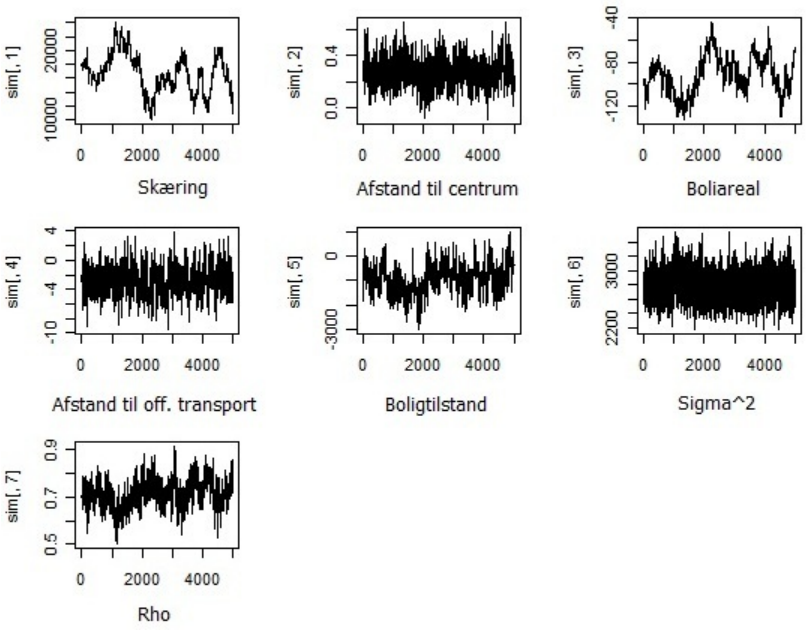
Vi vil nu sammenligne ovenstående resultater for modelparametrene med dem, der fås ved anvendelse af den bayesianske tilgangsvinkel til SAR-modellen. Vi bruger pakken *gibbs.met*, der kombinerer M-H-sampling og Gibbs-sampling, og vi definerer posteriorfordelingen $p(\boldsymbol{\beta}, \sigma^2, \rho | \mathcal{D})$ fra sætning 6.3 i kode 7.4. Matricen W er defineret som i kode 7.1.

```
install.packages("gibbs.met") #installerer pakke til algoritme med M−H i Gibbs−sampling
library(gibbs.met)

data1<-read.table("homegennemsnit.csv", sep=";", header=T) #datasæt med relevante baggrundsvariable til X−matricen

# Her laves en funktion, der udregner loglikelihoodfunktionen for den simultane posteriorfordeling
bayespostford <- function(x){
l = length(x) # antallet af parametre
beta = x[-c(l-1,l)] # parametre tildeles en plads i x
sigma = x[l-1]
rho = x[l]
if (sigma<0) return(-Inf) # Her sikres at sigma er positiv
if (rho<1/min(eigen(W)$values)) return(-Inf) # Her sikres at rho er mindre end den inverse af den mindste egenværdi for W, det ime = 1/min(eigen(W)$values)
if (rho>1) return(-Inf) # Her sikres at rho ikke er større end 1
a1 = a + n/2 # Hyperparametre i posteriorfordelingen defineres
Sigma1=solve(t(X)%*%X + solve(Sigma))
c1=Sigma1%*%(t(X)%*%(diag(94)-rho* W)%*%y + solve(Sigma)%*%c)
b1= b + (t(c)%*%solve(Sigma)%*%c+ t(y)%*%t(diag(94)-rho*
    W)%*%(diag(94)-rho* W)%*%y-t(c1)%*%solve(Sigma1)%*%c1)/2
return(-(a1 + k/2 + 1)*log(sigma^(2)) + log(det(diag(94)-rho* W)) -
    (1/(2*sigma^2)*(t(beta - c1)%*%solve(Sigma1)%*%(beta-c1)+2*b1))) #
    Posteriorfordelingen
}
a=0.01 # selvvalgte værdier for hyperparametre
b=0.01
c=c(0,0,0,0,0)
n=94 # antal kommuner
Sigma= 10^10*diag(5)
k=5 # antallet af parametre i betavektoren
y=data1[,13] # kvadratmeterprisen for boliger
```

```
28 X=cbind(1,as.matrix(data1[,c(6,4,9,10)])) # Matricen X laves med
       baggrundsvariablene: Afstand til centrum, boligareal, afstand til offentlig transport og boligtilstand
       fra datasættet
29
30 sim = gibbs_met(bayespostford, no_var=7,
       ini_value=c(18000,0.3,-98,-3,-1000, 2800,
       0.7),iters=5000,iters_met=2,
       stepsizes_met=c(2600,0.3,16,2.5,700,500,0.05)) # Simulation af parametre fra
       posteriorfordelingen
31 # no_var er antallet af komponenter i Gibbs-sampleren
32 # ini_value er vektor med begyndelsesværdier
33 # iters er antallet af iterationer
34 # stepsizes_met er standardafvigelserne i forslagsfordelingerne
35
36 plot(sim[,1],type="l") # traceplot
```

R-kode 7.4: Kode til at tilpasse en bayesiansk model til data ved hjælp af M-H i Gibbs.

Da vi ikke har betydelig a priori-viden, vælges de fleste hyperparameterværdier til at være tæt på nul, jævnfør overvejelser i afsnit 6.2, da det afspejler manglende information om parameterværdierne. Vi har hermed anvendt priorfordelinger, der ikke er informative. Kovariansmatricen *Sigma* får en relativt høj værdi på diagonalen for at angive stor usikkerhed på parametrene og ingen korrelation, idet det er en diagonalmatrix. Der vælges begyndelsesværdier tæt på maximum likelihood-estimaterne, der er vist på figur 7.5. Kode 7.4 giver blandt andet det traceplot, der er vist på figur 7.9.



Figur 7.9: Traceplot for estimaterne for parameteren ρ.

Traceplottet her viser, at estimaterne konvergerer mod en ligevægt, og hvis der foretages endnu flere iterationer, vil tendensen sandsynligvis være tydeligere. Desuden kan yderligere finjustering af standardafvigelserne i forslagsfordelingerne medføre, at der

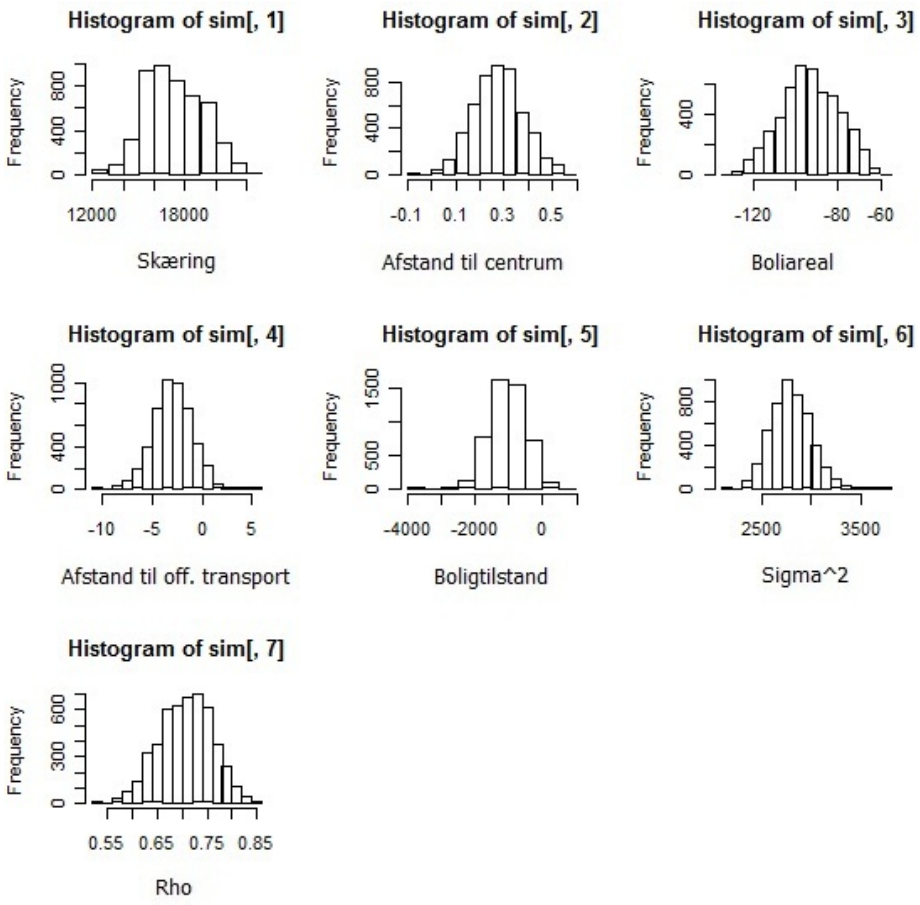
kan være mere tydelig konvergens. Det skal her bemærkes, at burn-in ikke er fjernet på traceplottet.

Kode 7.5 laver histogrammer og middelværdi for parametrene.

```
1 hist(sim[,1]) # approksimeret marginalfordeling for parameter
2
3 mean(sim[,1]) # approksimeret middelværdi for parameter(punktestimat)
```

R-kode 7.5: Kode til histogrammer og middelværdi for parametre.

Histogrammerne for parametrene ses på figur 7.10.



Figur 7.10: Histogram for ρ.

Der er her blevet anvendt M-H i Gibbs til at estimere modelparametrene $\boldsymbol{\beta}$, σ^2 og ρ, og ved brug af kode 7.5 fås estimaterne til

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} 17.296{,}84 \\ 0{,}2763 \\ -93{,}2495 \\ -3{,}0215 \\ -1032{,}26 \end{bmatrix}$$
$$\hat{\sigma}^2 = 7.807.319$$
$$\hat{\rho} = 0{,}7071.$$

Det ses at generelt er estimaterne her en del mindre end ved maximum likelihood-metoden, bortset fra afstand til offentlig transport og ρ, som er lidt højere end ved maximum likelihood-metoden, hvilket vil sige, at den rumlige afhængighed ved anvendelse af MCMC med M-H i Gibbs er større. Forskellen på estimaterne for de to metode kan ses som et resultat af, at der kun er anvendt 5000 simuleringer, og desuden kan der justeres yderligere på variansen. Derudover vil det give et mere retvisende resultat at tage burn-in fra, inden middelværdi og residualplot findes. Hvis vi havde haft en betydelig a priori-viden, kunne den være blevet inddraget og måske have givet mere præcise estimater, hvilket ikke er en mulighed ved maximum likelihood-metoden.

Vi kan udregne et CPI for parametrene, hvilket ses i kode 7.6.

```
1 simm1<-sort(sim[,1]) # vektoren med alle værdier for den første betaværdi sorteres
2
3 simm1[floor(5000*0.025)] # 2.5% fraktilen
4 simm1[floor(5000*0.975)] # 97.5% fraktilen
```

R-kode 7.6: Kode til at finde CPI.

Det ses, at CPI for parametrene $\boldsymbol{\beta}$, σ^2 og ρ er givet ved

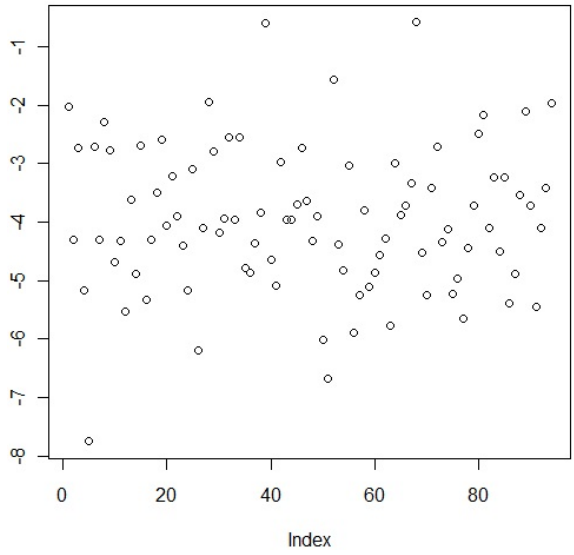
$$\boldsymbol{\beta}_1 \in [13.886{,}71; 20.889{,}74]$$
$$\boldsymbol{\beta}_2 \in [0{,}0825; 0{,}4784]$$
$$\boldsymbol{\beta}_3 \in [-120{,}1464; -67{,}6039]$$
$$\boldsymbol{\beta}_4 \in [-7{,}0781; 0{,}7455]$$
$$\boldsymbol{\beta}_5 \in [-2.131{,}21; -12{,}66]$$
$$\rho \in [0{,}6004; 0{,}8055],$$

og at intervallerne er smallere end ved maximum likelihood-metoden, hvilket indikerer større koncentration om middelværdien.

Residualplottet findes i kode 7.7 og ses på figur 7.11.

```
1 y=data1[,13] # Datavektor
2 X=cbind(1,as.matrix(data1[,c(6,4,9,10)])) # Designmatrix
3 plot((y - (solve(diag(94)-0.7071*W)%*%X%*%
4 c(17296.84,0.2763,-93.2495,-3.0215,1032.26)))/2794.16) # Residualplot
```

R-kode 7.7: Kode til at finde CPI.



Figur 7.11: Residualplot.

Residualplottet viser, at residualerne umiddelbart er jævnt fordelt og er tættere på nul end residualerne for maximum likelihood-metoden.

Vi kan hermed konkludere, at begge metoder taler for, at der er rumlig afhængighed mellem observationerne i datasættet fra HOME, og parameterestimaterne for hver metode tilnærmelsesvist ligner hinanden.

Konklusion 8

Vi har nu set på både klassisk og bayesiansk teori om modeller og modeltest inden for rumlige, økonometriske modeller med fokus på den rumligt afhængige SAR-model. Store dele af denne teori er blevet anvendt i praksis til analyse af et datasæt med boligpriser fra ejendomsmægleren HOME.

I indledningen introducerede vi en kort motivation for valg af emne samt et kort historisk overblik over indførelsen af rumlige økonometriske modeller, og hvordan rumlighed i modeller inddrages mere og mere i økonometriens verden. Derudover fremsatte vi en række overvejelser om metoden til rumlige modeller og rumlighed som begreb, og hvorledes den kan beskrives.

Herefter introduceredes SAR-modellen i detaljer med dens opbygning, parametre og egenskaber, og derudover så vi på gennemsnitlige påvirkninger for SAR-modellen, som kan anvendes til at sige noget om, hvordan ændringer i baggrundsvariable kan påvirke alle observationer. Ligeledes blev der gennemgået et eksempel på, hvordan man kan specificere vægtmatricen W, således at naborelationer kan inddrages i SAR-modellen.

Vi har set på maximum likelihood-estimation generelt til estimering af parametrene i SAR-modellen, og hvordan maximum likelihood-estimation giver en bedre estimering af SAR-modellens parametre, idet der tages højde for flere faktorer end ved brug af mindste kvadraters metode.

Monte Carlo-approksimation er blevet gennemgået med henblik på tilfælde, hvor der haves store datasæt, så approksimationer for logdeterminanten i SAR-modellen kræver særlig opmærksomhed. Herudover har vi analyseret, hvilke konsekvenser en sådan Monte Carlo-approksimation har for modelparameteren ρ, der har stor interesse i SAR-modellen. I den forbindelse har vi set på en anden metode til at beskrive gennemsnitlige påvirkninger for SAR-modellen, og der er blevet set på lukkede løsninger til SAR-modellen.

Vi har set på modeltest med henblik på SAR-modellens parametre, herunder hypotesetest samt begreberne signifikansniveau, p-værdi og konfidensinterval. Herudover har vi introduceret likelihood ratio-testen og Waldtesten, for hvilken der gøres nogle praktiske overvejelser. Desuden er begrebet residualer blevet introduceret, da sådanne ofte anvendes i forbindelse med grafisk undersøgelse af fejl i modellen ved modeltest.

For at sammenligne den klassiske tilgang med den bayesianske tilgang til SAR-modellen introducerede vi den bayesianske teori generelt. Den bayesianske metode er blevet gennemgået med hensyn til estimering af SAR-modellens parametre. Der har i den forbindelse været fokus på Markov Chain Monte Carlo-estimation anvendt i form af Markov Chain Monte Carlo-algoritmen med Gibbs-sampling og Metropolis-Hastings-

sampling. Der er kort blevet redegjort for gennemsnitlige påvirkninger i den bayesianske sammenhæng og begrebet CPI er blevet gennemgået.

Slutteligt anvendte vi store dele af ovenstående teori i praksis i programmet R til analyse af datasættet fra ejendomsmægleren HOME indeholdende en række oplysninger om boligsalg. Vi gjorde os en række overvejelser om, hvilke baggrundsvariable der skulle inddrages i modellen, samt hvad de forskellige størrelser i SAR-modellen udtrykte i forhold til vores data, og vi brugte Voronoi-diagrammet til at beskrive naborelationer mellem kommuner i SAR-modellen. Vi har estimeret SAR-modellens parametre samt analyseret deres betydning, og vi har set på en række test for disse. Desuden så vi på konfidensintervaller og residualplot for estimeringerne for modelparametrene, og vi så på de gennemsnitlige påvirkninger. Ligeledes undersøgte vi, hvordan vi ved at bruge priorfordelinger, der ikke er informative, kunne få estimeringer af SAR-modellens parametre tilsvarende estimaterne ved anvendelse af maximum likelihood. Vi så på traceplots for M-H-sampling i Gibbs, lavede residualplot og undersøgte kort histogrammer og CPI'er for alle parametrene.

I databehandlingsafsnittet fremkom det resultat, at estimaterne ved den bayesianske fremgangsmåde er en del mindre end ved maximum likelihood-metoden, bortset fra afstand til offentlig transport og ρ, som var lidt højere. Dette resultat indikerer, at den rumlige afhængighed ved den bayesianske metode er større end maximum likelihood-metoden, og at begge metoder underbygger en konklusion om rumlig afhængighed mellem observationerne. Parameterestimaterne for hver metode ligner tilnærmelsesvist hinanden, hvor der dog er flere faktorer, der kan have påvirket forskellen. For at få mere ens og dermed muligvis mere præcise resultater kan der foretages flere simuleringer i MCMC, burn-in kan tages fra, der kan tilføjes mere a priori-viden, eller der kan justeres yderligere på variansen.

CPI for parametrene $\boldsymbol{\beta}$, σ^2 og ρ i MCMC har smallere intervaller end ved maximum likelihood-metoden, hvilket indikerer større koncentration om middelværdien og dermed en smule mere præcise estimater. Begge residualplot er jævnt fordelte, men det bayesianske ligger tættest på nul og dermed kan understøtte en teori om mere præcise resultater ved denne metode.

Litteratur

Luc Anselin. *Thirty years of spatial econometrics.* Blackwell Publishing, 2010.

Portal Danmark A/S. *rejs.dk.* URL: http://www.rejs.dk/regioner.asp. Dato: 16-12-2012.

Sheldon Axler. *Linear Algebra Done Right.* Springer, 1997. ISBN ISBN: 978-0-387-98258-8.

Adelchi Azzalini. *Statistical Inference - Based on the likelihood.* Chapman & Hall/CRC, 1996. ISBN ISBN: 978 0412606502.

Norman R. Draper and Harry Smith. *Applied Regression Analysis.* 1998.

Wikipedia The Free Encyclopedia. *Determinant.* URL: http://en.wikipedia.org/wiki/Determinant#Derivative, a. Dato: 3-12-2012.

Wikipedia The Free Encyclopedia. *Invertible matrix.* URL: http://en.wikipedia.org/wiki/Invertible_matrix, b. Dato: 3-12-2012.

Wikipedia The Free Encyclopedia. *Matrix Calculus.* URL: http://en.wikipedia.org/wiki/Matrix_calculus, c. Dato: 3-12-2012.

Wikipedia The Free Encyclopedia. *Errors and residuals in statistics.* URL: http://en.wikipedia.org/wiki/Errors_and_residuals_in_statistics, d. Dato: 6-12-2012.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis.* Chapman & Hall, 1995. ISBN ISBN: 0 412 03991 5.

Zsusanna Horvath and Ryan Johnston. *AR(1) Time Series Process.* URL: http://www.math.utah.edu/~zhorvath/ar1.pdf, Dato 19/11-2011 .

Kaj Jensen. *Optocleaner.* URL: http://www.optocleaner.com/danmarks_kommuner.htm. Dato: 18-12-2012.

Krak. Krak. Dato: 05-12-2012.

Perko Lawrance. *Differential Equation and Dynamical Systems.* Springer, 2001.

Peter M. Lee. *Bayesian Statistics.* Wiley, 2004. ISBN ISBN: 978 0 470 68920 2.

James P. LeSage and Robert Kelley Pace. *Introduction to Spatial Econometrics.* CRC Press, Chapman & Hall, 2009. ISBN ISBN: 978 1 4200 6424 7.

Peter Olofsson. *Probability, Statistics, and Stochastic Processes.* Wiley, 2005. ISBN ISBN: 978 0471679691.

Kenneth H. Rosen. *Discrete Mathematics and its Applications.* McGraw-Hill, 2007.

Danmarks Statistik. Statistikbanken. Dato: 05-12-2012.

Melanie M. Wall. *A close look at the spatial structure implied by the CAR and SAR models.* Elsevier B.V., 2003.

Marc Wick. *GeoNames.* URL: `http://www.geonames.org/postal-codes/postal-codes-denmark.html`. Dato: 29-11-2012.

Appendiks A

A.1 Tidskompleksitet

Dette afsnit er baseret på [Rosen, 2007, kapitel 3].

I praksis anvendes ofte algoritmer til at behandle svære udtryk, der indeholder kompliceret matrixregning med store matricer. Derfor er det praktisk at kunne sammenligne hastigheden og dermed effektiviteten af algoritmer anvendt med forskellige regnemetoder. Udtrykket *tidskompleksitet* beskriver, hvor mange operationer en algoritme bruger for at løse et givet problem. Sådanne operationer kan være forskellige basisoperationer som multiplikation eller udbytning af tal. Tidskompleksiteten er et udtryk for, hvor meget arbejdskraft computeren skal bruge, og den angiver ikke det reelle tidsforbrug. Når man udtrykker algoritmers effektivitet gennem tidskompleksitet, ses der bort fra, hvilken software og hardware der anvendes, og der regnes med, at alle operationer tager lige lang tid at udføre, hvilket selvfølgelig er en forenkling af praksis. Tidskompleksitet kan ses som et estimat af sværhedsgraden for udførelsen af en algoritme baseret på antallet operationer, som en algoritme skal bruge i takt med, at inputtet n vokser. Sværhedsgraden beskrives ved Store-O-notationen, der er udtrykt i følgende definition.

Definition A.1 (Store-O):
Lad f og g være reelle funktioner defineret for en mængde bestående af heltal eller reelle tal, så $x \in \mathbb{R}$ eller $x \in \mathbb{Z}$. Funktionen $f(x)$ er $O(g(x))$, hvis der findes konstanter C og k, så

$$\left|f(x)\right| \leq C \cdot \left|g(x)\right| \quad \text{når } x > k.$$

Typiske tidskompleksiteter er vist i følgende tabel.

Kompleksitet	Terminologi
$O(1)$	konstant kompleksitet
$O(\ln(n))$	logaritmisk kompleksitet
$O(n)$	lineær kompleksitet
$O(n\ln(n))$	$n\ln(n)$ kompleksitet
$O(n^k)$	polynomisk kompleksitet
$O(k^n)$	eksponentiel kompleksitet
$O(n!)$	faktoriel kompleksitet

Appendiks B

B.1 Normalfordelingen

Definition B.1 (Normalfordeling):
En stokastisk vektor $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$ siges at være n-dimensionalt normalfordelt med parametrene $\boldsymbol{\mu}$ og Ω, hvis den har den simultane tæthedsfunktionen

$$p(\mathbf{y}) = \frac{1}{(2\pi)^{n/2}|\Omega|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{y}-\boldsymbol{\mu})^T \Omega^{-1}(\mathbf{y}-\boldsymbol{\mu})\right\},$$

hvor $\mathbf{y} = (y_1, y_2, \ldots, y_n)^T \in \mathbb{R}^n$, $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)^T \in \mathbb{R}^n$ er middelværdivektor, og hvor Ω er en positivt semidefinit $(n \times n)$-matrix. Dette noteres som

$$\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \Omega).$$

Man kan vise, at $\boldsymbol{\mu}$ og Σ er henholdsvist middelværdivektoren og kovariansmatricen udtrykt ved

$$\boldsymbol{\mu} = \begin{bmatrix} \mathrm{E}[Y_1] \\ \vdots \\ \mathrm{E}[Y_n] \end{bmatrix} \quad \text{og} \quad \Omega = \begin{bmatrix} \mathrm{Cov}[Y_1, Y_1] & \cdots & \mathrm{Cov}[Y_1, Y_n] \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}[Y_n, Y_1] & \cdots & \mathrm{Cov}[Y_n, Y_n] \end{bmatrix}.$$

Dette kan ses i [Lee, 2004, s. 304].

Desuden gælder der, at marginalfordelinger, linearkombinationer og betingede fordelinger også er normalfordelte.

B.2 χ^2-fordeling

Definition B.2:
En stokastisk variabel $Y \in \mathbb{R}$ siges at være χ^2-fordelt med ν frihedsgrader, hvis den har tæthedsfunktionen

$$p(y) = \frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)} y^{\frac{\nu}{2}-1} \exp\left\{-\frac{1}{2}y\right\} \qquad 0 < y < \infty,$$

hvor $\nu \in \mathbb{Z}_+$. Dette noteres

$$Y \sim \chi^2_\nu.$$

Dette kan ses i [Lee, 2004, s. 289-290].

B.3 Den uniforme fordeling

Definition B.3 (Uniform fordeling):
En stokastisk variabel $Y \in \mathbb{R}$ siges at være uniformt fordelt på intervallet $[a,b]$, hvis den har tæthedsfunktionen

$$p(y) = \frac{1}{b-a} \quad a \leq y \leq b.$$

Dette noteres som

$$Y \sim U(a,b).$$

Middelværdien og variansen er givet ved

$$\mathrm{E}[Y] = \frac{a+b}{2}$$
$$\mathrm{Var}[Y] = \frac{(b-a)^2}{12}.$$

Dette kan ses i [Lee, 2004, s. 296-297].

B.4 Invers gammafordeling

Definition B.4:
En stokastisk variabel $Y \in \mathbb{R}$ siges at være invers gammafordelt med parametrene a og b, hvis den har tæthedsfunktionen

$$p(y) = \frac{b^a}{\Gamma(a)} y^{-(a+1)} \exp\left\{-\frac{b}{y}\right\} \qquad 0 < y < \infty,$$

hvor Γ er gammafunktionen givet ved $\Gamma(a) = (a-1)!$ for positive heltal, og $a, b \in \mathbb{R}_+$. Dette noteres

$$Y \sim IG(a,b).$$

Gammafunktionen er defineret for positive heltal og alle komplekse tal, men ikke for negative heltal og nul. For komplekse tal med positiv realdel er den defineret ved det uegentlige integrale

$$\Gamma(n) = \int_0^\infty e^{-t} t^{n-1} dt.$$

Middelværdi og varians for den inverse gammafordeling er givet ved

$$\mathrm{E}[Y] = \frac{b}{a-1} \qquad \text{givet } a > 1$$
$$\mathrm{Var}[Y] = \frac{b^2}{(a-1)^2(a-2)} \qquad \text{givet } a > 2.$$

Dette kan ses i [Gelman et al., 1995, s. 474].

Det ses, at der er en tæt forbindelse mellem den inverse gammafordeling og χ^{-2}-fordelingen. Hvis $Y \sim IG(a,b)$ med $a = \frac{v}{2}$ og $b = \frac{1}{2}$, så får vi netop fordelingen for χ^{-2}-fordelingen med v frihedsgrader.

B.5 Betafordeling

Definition B.5:
En stokastisk variabel $Y \in \mathbb{R}$ siges at være betafordelt med parametrene a og b, hvis den har tæthedsfunktionen

$$p(y) = \frac{1}{B(a,b)} Y^{a-1}(1-Y)^{b-1} \qquad 0 < y < 1,$$

hvor $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} = \int_0^1 t^{a-1}(1-t)^{b-1}\,dt$ og $\Gamma(n) = (n-1)!$ for positive heltal. Dette noteres

$$Y \sim B(a,b).$$

Middelværdi og varians for en betafordeling er givet ved

$$\mathrm{E}[Y] = \frac{a}{a+b}$$
$$\mathrm{Var}[Y] = \frac{ab}{(a+b)^2(a+b+1)}.$$

Dette kan ses i [Lee, 2004, s. 292-293].

Desuden skal det bemærkes, at for $a = b = 1$ haves en standard uniform fordeling $U(0,1)$. Betafordelingen ligger som udgangspunkt i intervallet $[0,1]$, men hvis den skaleres, kan den komme til at ligge i andre intervaller, hvis det ønskes. Når $a = b = 0$ haves en fordeling, der er proportional med $p^{-1}(p-1)^{-1}$, og dermed repræsenteres en fuldstændig usikkerhed.

Appendiks C

C.1 Matrixregneregler

Regnereglerne er at finde i [Encyclopedia, a], [Encyclopedia, b] og [Encyclopedia, c].

Regneregel C.1:
For matricen $A(a)$, hvor a er et reelt tal, gælder, at

$$\begin{aligned}\frac{\partial |A|}{\partial a} &= |A| \operatorname{tr}\left(A^{-1}\frac{\partial A}{\partial a}\right) \Rightarrow \\ \frac{\partial \left|I_n - \rho W\right|}{\partial \rho} &= \left|I_n - \rho W\right| \operatorname{tr}\left(\left(I_n - \rho W\right)^{-1}\frac{\partial \left(I_n - \rho W\right)}{\partial \rho}\right) \\ &= \left|I_n - \rho W\right| \operatorname{tr}\left(-\left(I_n - \rho W\right)^{-1} W\right).\end{aligned}$$

Regneregel C.2:
For en matrix A gælder $I_n = AA^{-1}$, og det haves derfor, at

$$\begin{aligned}0 = \frac{\partial}{\partial a} I_n = \frac{\partial}{\partial a}\left(AA^{-1}\right) = \frac{\partial A}{\partial a}A^{-1} + A\frac{\partial A^{-1}}{\partial a} \Rightarrow \\ \frac{\partial A^{-1}}{\partial a} = -A^{-1}\frac{\partial A}{\partial a}A^{-1}.\end{aligned}$$

Regneregel C.3:
For matricer A, B, C og D gælder det, at

$$\operatorname{tr}(ABCD) = \operatorname{tr}(DABC) = \operatorname{tr}(CDAB) = \operatorname{tr}(BCDA).$$

Regneregel C.4:
For en symmetrisk matrix A og en vektor $\boldsymbol{\beta}$ gælder det, at

$$\frac{\partial}{\partial \boldsymbol{\beta}}\left(\boldsymbol{\beta}^T A \boldsymbol{\beta}\right) = 2\boldsymbol{\beta}^T A.$$

Regneregel C.5:
For en matrix A og en vektor $\boldsymbol{\beta}$ gælder det, at

$$\frac{\partial}{\partial \boldsymbol{\beta}^T}\left(\boldsymbol{\beta}^T A\right) = A.$$